Infinitude of the Prime Numbers ❖ From Matter
to Life : Chemistry?! ❖ Intel's New
P6 Processor ❖ The Earth's Changing
Climate ❖ Diversity of Bats

Editorial

N Mukunda, Chief Editor

Many readers have written to us with words of appreciation after having seen the initial issues of **Resonance**. They have also made several valuable suggestions for improvements. We are grateful for the former, and will pay attention to the latter as we go along. Slipping into the jargon of the times, we want to make **Resonance** 'reader-friendly' and 'author-friendly', neither stiff nor dry like a conventional research journal in science. An informality of tone in writing, while being accurate in content — that is what we would like to achieve.

Jayant Narlikar's article in his series on the "Origin (?) of the Universe" moves up to the Big Bang. Soon after Einstein formulated his field equations of general relativity in 1915, he modified

AYAN GUHA

Oh God! Its expanding !!!

After Hubble's discovery in 1929 of the expansion of the universe, Einstein ended up calling his earlier modification "the biggest blunder of my life!"
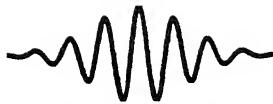
them in 1917 by adding the so-called *Cosmological Constant* term. This was in a bid to avoid solutions which described an expanding universe — in 1917 such solutions appeared to Einstein to be unphysical. But with Hubble's discovery in 1929 of the expansion of the universe, the situation was reversed; Einstein ended up calling his earlier modification "the biggest blunder of my life!"

Slipping into the jargon of the times, we want to make Resonance *'reader-friendly' and 'author-friendly', neither stiff nor dry like a conventional research journal in science.*

Jean -Marie Lehn, one of three Nobel Prize winners in Chemistry in 1987, is essentially the creator of the new field of supramolecular chemistry. He has visited India several times, most recently in December 1995 to deliver the Rajiv Gandhi Science and Technology Lecture at Hyderabad. We are thankful to C N R Rao for making it possible for *Resonance* to publish the text of this lecture in its entirety. Lehn highlights the stupendous qualitative leaps that chemistry has taken over the past century and a half — from a mechanistic view of structure and function, the "lock and key" concept of Emil Fischer, to a much deeper view encompassing the ability of molecular structures to capture, create and handle information. In this sense is the bridge to biology established. At the same time, as Lehn emphasizes, while biology scores over chemistry in complexity, the latter comes into its own in the diversity of the materials and structures that can be created.

In describing Fermat's work in the January 1996 issue (inside back cover), in a burst of enthusiasm we described his principle of least time in optics as the earliest example of a minimum principle in physics. This was a fortunate error, as it inspired Gangan Prathap to correct us and so tell us about the "Alexandrian Hero" in this issue!

Fermat and the Minimum Principle

Gangan Prathap

The inside back cover of the January 1996 issue of *Resonance* describes Pierre-Simon de Fermat (1601-1665) as "the discoverer of the principle of least time in optics, the earliest example of a minimum principle in physics."

Arguably, least action and minimum principles were offered or applied much earlier. This (or these) principle(s) is/are among the fundamental, basic, unifying or organizing ones used to describe a variety of natural phenomena. It considers the amount of energy expended in performing a given action to be the least required for its execution. Again, where motion is concerned, a particle or wave chooses the shortest possible path in moving from one point to another (as in the law of reflection, which goes back to Hero), or tries to complete the motion in the shortest possible time (as in the law of refraction, which Fermat established).

Engineers believe that Archimedes, who lived during the 200's BC (nineteen centuries before Fermat), derived the principle of the lever using an approach that is essentially a variation of the least action principle. This technique, now known as the virtual work principle, is used as the basis for structural mechanics, the science underlying structural engineering.

Even in optics, as already alluded to earlier, we find a famous application of this law of economy of physical behaviour to optics being made by Hero (or Heron) of Alexandria who lived during the first century AD. He discovered the law of reflection which every school child is familiar with. This law states that angles of incidence and reflection formed by a light ray incident on a plane mirror are determined by the condition that the ray travel from its source to its reflected position along the shortest possible path. This was some sixteen centuries before Fermat demonstrated that the law of refraction of light also follows from the minimum principle.

> Engineers believe that Archimedes, who lived during the 200's BC, derived the principle of the lever using an approach that is essentially a variation of the least action principle.

Suggested Reading

J R Newman. The World of Mathematics, Simon and Schuster, 1956.
R P Feynman, R B Leighton, M Sands. The Feynman Lectures on Physics, Vol.1 & 2. Addison-Wesley, Reading, Mass. 1963 and Narosa, New Delhi, 1987, pp.315-316 and 956-974.

Front Cover
A space-filling model of a self-assembly of five ligands of two different types and six Cu(I) ions spontaneously producing a cage structure. See the article by J. -M. Lehn for details.

Back Cover
R B Woodward (1917 - 1979).
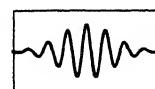
Origin (?) of the Universe

3. The Big Bang

Jayant V Narlikar

In this part of the series we look at the simplest cosmological models based on the simplifying assumptions of the Weyl postulate and the cosmological principle. These models were discovered independently by Friedmann, Lemaitre and Robertson in the 1920s. They led to the striking conclusion that the universe started in an enormous explosion often called the Big Bang.

Relativistic Cosmology

The simplifying postulates described in part 2 of the series (*Resonance*, Vol. 1, No. 2) allow the theoretician to construct mathematical models of the expanding universe. In 1922 the Russian physicist Alexander Friedmann made such attempts and arrived at what are today known as the *Friedmann models*. During 1922-24 when Friedmann was constructing these models he needed a theory of gravity to determine the large scale dynamics of the universe. Why a theory of gravity?

Physicists today know of essentially four basic interactions in nature: the strong and the weak interactions which describe the microscopic behaviour of subatomic particles, the electromagnetic interaction which describes the force between electric charges, at rest or in motion and the gravitational interaction which is none other than that first quantified by Isaac Newton three centuries ago.

In the 1920s the knowledge of the first two of these four interactions was rudimentary. But it was clear that these microscopic forces were of very short range, being confined to atomic nuclei and would not be of much relevance to the large scale structure of

Jayant Narlikar, Director, Inter-University Centre for Astronomy and Astrophysics, works on action at a distance in physics, new theories of gravitation and new models of the universe. He has made strong efforts to promote teaching and research in astronomy in the universities and also writes extensively in English and Marathi for a wider audience on science and other topics.

the universe today. So also the electromagnetic interaction which by then was well understood at the classical level. Although it is of long range (i.e. its influence can be felt over arbitrarily large distances albeit with decreasing intensity), in order for it to be effective in cosmology, the major constituents of the universe would have to be electrically charged. The observations indicate (as they did in the 1920s) that matter on the large scale is electrically neutral, and so this force is also ruled out of contention.

This left gravity — another long range force which though very weak at the atomic level, comes into its own at the cosmological scales of large masses as in the case of galaxies and clusters of galaxies. However, it is here that the Newtonian picture becomes suspect. For, it involves the concept of instantaneous gravitational action at a distance which is inconsistent with special relativity. Newtonian dynamics also needs to be revised so as to be consistent with special relativity. By 1922, Einstein's general theory of relativity was already being established as a theory that was free from the above two conceptual difficulties and in better agreement with the solar system observations. Thus it was natural for Friedmann to use the framework of general relativity to describe the simplest cosmological models.

The Einstein Universe

Relativistic cosmology, as this subject is now known, started even earlier than Friedmann's pioneering work. In 1917, Einstein himself had attempted to obtain a simple relativistic model of the universe; but he had made the assumption that the universe was a static system. At that time there was no conclusive evidence for the expanding universe (Hubble's result came in 1929) and so, in looking for a simple model, Einstein was perfectly right in assuming that there is no large scale motion in the universe.

However, Einstein soon discovered that there is no static solution to his relativistic equations! Rather than look for a dynamical solution like Friedmann did five years later, Einstein sought to

Physicists today know of essentially four basic interactions in nature: the strong and the weak interactions which describe the microscopic behaviour of subatomic particles, the electromagnetic interaction which describes the force between electric charges, at rest or in motion, and the gravitational interaction which is none other than that first quantified by Isaac Newton three centuries ago.

modify his field equations so that he could get a *static* solution. For this he had to introduce a 'cosmological term' describing an extra cosmic force of repulsion. The rationale for Einstein's approach can be understood by the following simple argument.

Imagine two masses m and M separated by a distance r in the Newtonian framework. If the two masses are left to themselves they cannot remain at rest but will move towards each other with an attractive force GmM/r^2. To have a static situation we can introduce a repulsive force λr which will balance the above force if the distance r is adjusted to be

$$r = (GmM / \lambda)^{1/3}.$$

Here λ is a constant. In Einstein's general relativistic formulation the constant λ appears in the same spirit but the mathematical details are different. Einstein called it the *cosmological constant*. Its effect on the whole universe was to adjust its radius so that the repulsive force exactly balances the gravitational contraction of the universe. The radius of the universe is then given by the formula

$$R=[2GM/(\pi c^2\lambda)]^{1/3}$$

where M is the mass of the universe. (Note: the dimension of λ is different from that in the Newtonian example.)

What do we mean by the 'mass' of the universe? How can an infinite system have a finite mass? The answer to this query is that the universe is unbounded but has a finite volume. How can space be unbounded but finite? The answer is that in Einstein's model, the universe is the surface of a hypersphere in four dimensions. Just as the two dimensional surface of an ordinary sphere of radius R is finite and has an area equal to $4\pi R^2$, the three dimensional surface of a hypersphere is finite and has a finite volume equal to $2\pi^2R^3$. In both the examples, there is no boundary to the space, i.e., it is unbounded. Thus it is perfectly legitimate to talk of an unbounded but finite system.

Einstein's model, however, did not remain in contention once

Hubble's discovery became well established. Einstein himself withdrew it along with the cosmological term saying that it was his greatest blunder! Others may disagree. For example, in 1917 W de Sitter wrote a paper in which he solved Einstein's equations with the λ-term, and obtained a model in which the universe expands with a scale factor that increases exponentially with time. The *Hubble constant* for this universe is given by

$$H = [(1/3) \lambda c^2]^{1/2}.$$

Thus this model is apparently closer to reality than Einstein's because it expands. But it isn't, because this universe is empty! Nevertheless, as we shall see, the de Sitter universe has played an important role in cosmology in different contexts. We will return to the λ-term later in the series.

The Simplest Friedmann Models

During the early twenties, however, the notion of the expanding universe was not established and thus Friedmann's models remained of academic interest and were not widely known; even Einstein did not show any enthusiasm for them. Later in 1927, G Lemaitre (a priest!) and H P Robertson independently worked out similar models. And in the mid-1930's Robertson and A G Walker independently produced a mathematically rigorous derivation of the geometrical features of space and time starting from the symmetries of the Weyl Postulate and the Cosmological Principle (see part 2 of the series). For this reason these spacetimes are called the *Robertson-Walker* (R-W) spacetimes.

Figure 1 Surfaces of different curvatures: The figure illustrates the idea of curvature for two dimensional surfaces. The flat surface has zero curvature, the sphere has positive curvature while the saddle shaped surface has negative curvature. If triangles are drawn on the three surfaces their internal angles will add up to 180°, more than 180° and less than 180° respectively.
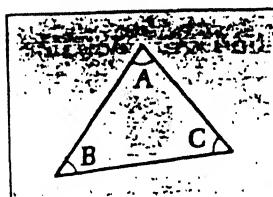
Zero Curvature	Positive Curvature	Negative Curvature
A + B + C = 180°	A + B + C > 180°	A + B + C < 180°
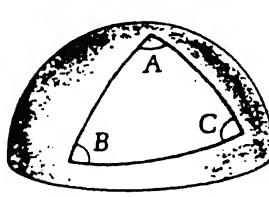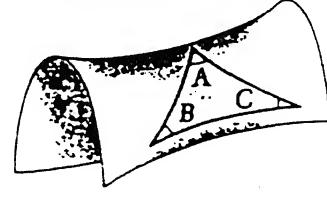
The three-dimensional spaces described by the sections at cosmic time $t = $ constant in the R-W spacetimes can be characterized by a single parameter k which specifies what type of curvature these spaces have. For $k=0$ we have spaces of zero curvature which have the *flat* Euclidean geometry that we are so familiar with in everyday life. For $k=1$, the spaces have positive curvature and these are similar to the space in Einstein's universe. These spaces are called *closed* and they have the feature that if we continue in a straight line in any direction, we would return eventually to the starting point. (Analogy: going in a 'straight' line on the earth would bring us back to where we started). The spaces given by $k=-1$, likewise have negative curvature and describe an *open* universe. (Analogy: the surface of a horse saddle.) The parameter k is therefore called the *curvature parameter*.

The simplest solutions obtained by Friedmann had all three possibilities for the curvature parameter. Einstein's equations relate this to the contents of the universe. But what did the universe contain by way of matter? The simplest models had matter in the form of 'dust', that is, pressureless fluid. This is an idealization of the real universe which has small random motions of galaxies and hence very small cosmic pressure. We will return to this aspect in part 4 of this series.

The formidable Einstein equations reduce under these conditions to just two relations which describe the linear scale factor $S(t)$ as a function of time in terms of the curvature parameter k and the density ρ of the universe. These relations are given below:

$$S^3 \rho = \text{constant}$$

$$[(dS/dt)^2 + kc^2] / S^2 = 8\pi G\rho / 3.$$

The first relation is simply the law of conservation of matter. It tells us that as the linear size of space expands in proportion to S, its density falls as the inverse cube of S. The second relation is a dynamical one which tells us that the rate of expansion is made slower by the gravitational attraction within the universe.

If we solve these relations together, we get $S(t)$ as a known function which, in the special case of the flat universe would correspond to $t^{2/3}$. The following features are noticed in general. All three types of solutions have similar beginnings. The universe 'begins' its existence at the stage when $S=0$. It expands in an explosive fashion $(\mathrm{d}S/\mathrm{d}t \longrightarrow \infty \text{ as } S \longrightarrow 0)$ at an epoch a finite time ago. Thereafter it either continues to expand till S becomes infinite $(k = 0,-1)$ or its expansion comes to a halt and it contracts back to $S = 0$ (for $k = 1$).

Some Cosmological Parameters

How do these models relate to Hubble's law? A simple analogy with a laboratory system will illustrate how the Hubble constant can be obtained from the expanding Robertson-Walker models.

Suppose you heat a metal ruler. It will expand. Let the function $S(t)$ describe how the original length l of the ruler grows with time. Keeping $S(t)=1$ at $t=0$ we say that the expanding length of the ruler is $lS(t)$. Thus, when seen from one end of the ruler, the other end appears to move with velocity $l\mathrm{d}S/\mathrm{d}t$, whereas the distance between the two ends is $lS(t)$. Hence velocity divided by distance is $(\mathrm{d}S/\mathrm{d}t)/S$, which is the 'Hubble constant'.

Apply this analogy to the cosmic ruler between two galaxies and

What did the universe contain by way of matter? The simplest models had matter in the form of 'dust', that is, pressureless fluid. This is an idealization of the real universe which has small random motions of galaxies and hence very small cosmic pressure.
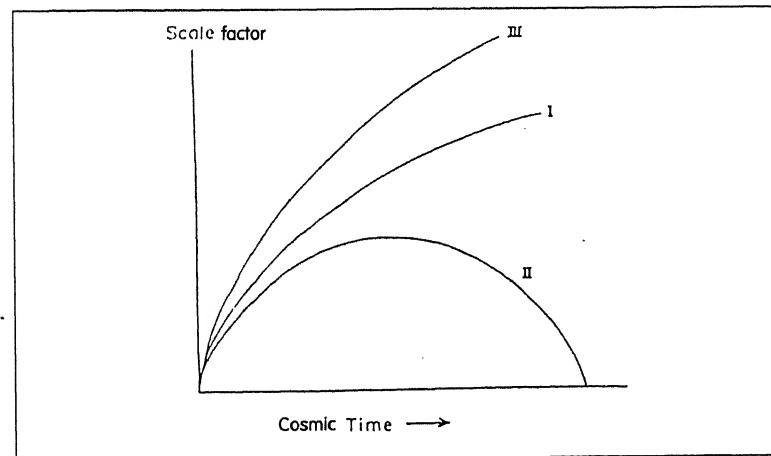
Figure 2 *Three simple Friedmann models: The figure shows schematically how the scale factor changes with cosmic time in the three different Friedmann models. Curve I corresponds to the flat case (k=0), curve II to the closed case (k=1) and curve III to the open case (k=-1). All three curves meet where the axes intersect. This point, identified with the Big Bang event has the scale factor equal to zero. We may consider the cosmic clock to start ticking from this epoch.*

you get the observed Hubble constant as

$$H = (d\,S/d\,t)\,/\,S.$$

A rigorous analysis based on general relativity confirms this simple minded derivation. The analysis also gives the *redshift z* as

$$1 + z = S(t_0)\,/\,S(t_1).$$

Here the redshift z is the fractional increase in the wavelength of light from the observed galaxy. It is assumed that light left the galaxy at the earlier epoch t_1 in order to reach us at the present epoch t_0. Since the observations show that $z > 0$, we conclude that the scale factor of the universe has increased between t_1 and t_0. That is, the universe is expanding.

The three types of Friedmann models are also distinguished by the *density parameter* which is usually denoted by Ω. If we take the flat ($k=0$) model, then calculations tell us that the cosmic density ρ_c is related to the Hubble constant by a simple formula:

$$\rho_c = 3H^2 \,/\, (8\pi G).$$

We then define the density of any other model by the formula

$$\rho = \Omega\,\rho_c$$

Again, detailed calculations show that for the closed ($k=1$) models $\Omega > 1$, whereas for the open ($k= -1$) models we have $\Omega < 1$, We will refer to ρ_c as the *critical density* or the *closure density*.

The Big Bang

The past epoch when S was equal to zero is called the epoch of the Big Bang. At this epoch the density is infinite and so is the curvature of spacetime. In fact, standard equations of physics break down and physicists encounter what they refer to as a *singularity*. Clearly, there is no justification for pushing our mathematical model further back in time beyond the Big Bang epoch. Cosmologists like to identify this epoch with the origin of the universe. In the next part of the series we will examine the state of the universe at epochs very close to the Big Bang.

Suggested reading

J V Narlikar. The Primeval Universe. Oxford University Press. 1988.

H Bondi. Cosmology. Cambridge University Press. 1960.

S Weinberg. Gravitation and Cosmology. Wiley. 1972.
Very technical but a classic.

Address for correspondence
Jayant V Narlikar
IUCAA, PB No. 4,
Ganeshkhind,
Pune 411 007, India

Algorithms

2. The While-Construct

R K Shyamasundar

In this article, we consolidate the introductory concepts developed in the previous article of the series (*Resonance*, Vol.1, No.1) and develop the *iterative construct*, one of the most important control constructs used to describe algorithms.

In the last article, we learnt about *assignment* and other basic commands which are imperative commands to a processor. Further, we discussed the basic control structures which include sequential composition and the *test* (or more specifically, the *if-then-else*) construct. Using these constructs, we developed the basic flowchart language for describing algorithms. In this article, we continue the discussion of control constructs and their representation using flowcharts. We describe the 'while-construct' which is one of the most widely used iterative control constructs for describing algorithms.

R K Shyamasundar
is Professor of
Computer Science at
TIFR, Bombay
who has done extensive
research in various
foundation areas of
computer science.

Iteration

We concluded the last article with the question: "Is it possible to obtain a concise flowchart to find the sum of the first N natural numbers?" We hinted that it was possible, by using a construct in which the number of times a set of commands is executed depends on the values of certain variables. Such a construct, referred to as the 'while-construct', is shown in *Figure 1*. The construct is interpreted as follows: Test for B; if the test leads to the answer "NO", then we have reached the end; otherwise the control goes to block S, after which the process repeats.

It is important to note that B is false (usually denoted by $\neg$ B where $\neg$ denotes the logical negation unary operator) on termination of the

Figure 1 *An important algorithmic construct called 'while loop'.*

while-loop. The textual representation of the construct is:

$$while\ B\ do\ S\ endwhile$$

When these operations terminate, we can assert that $\neg\,B$ (i.e., complement of B) holds.

Example 1: Going back to summing the first N natural numbers.

We describe an algorithm that can be used for any N. The idea is to keep track of the numbers we have already added and exit when we have added all the N numbers. The flowchart shown in *Figure 2* describes such an algorithm. We see that the same algorithm works for any value of N (fixed a priori). The textual algorithm (referred to as 'code') corresponding to the flowchart is given in *Table 1*. This algorithm solves the problem of adding the first N natural numbers for any value of N. We may add the box 'read N', shown in *Figure 3*, to the top of the flowchart given in *Figure 2*. It accomplishes the task of substituting the value of N in the flowchart of the given program. In other words, when *read N* is executed, the variable N takes the value from the given input.

Example 2: Euclid's Algorithm..

We now describe Euclid's algorithm for computing the greatest common divisor *(gcd)* of two positive integers m and n. By *gcd*, we mean the largest positive number that exactly divides both m and n. A naive way of obtaining the *gcd* is to factor both the numbers and take the common factors. However, such a scheme is quite tedious.

The Greek philosopher Euclid provided a better solution to this problem. Let us see the reasoning behind Euclid's algorithm. Let x be equal to *gcd* (m,n) where $m \geq n$. Then, we observe the following:

- $x \leq n$ since $n \leq m$. That is, the maximum value of x is bounded by the smaller of the two numbers (i.e., by n).
- $x = n$ implies that n exactly divides m.
- From the definition of *gcd*, we see that *gcd* $(m,n) =$ *gcd* (n,m).

Euclid's Algorithm

We learn to compute the greatest common divisor *(gcd)* in our primary school arithmetic classes. Although popularly known as Euclid's Algorithm, it was described by Euclid's predecessor Eudoxous. The ancient Chinese had also discovered this algorithm.
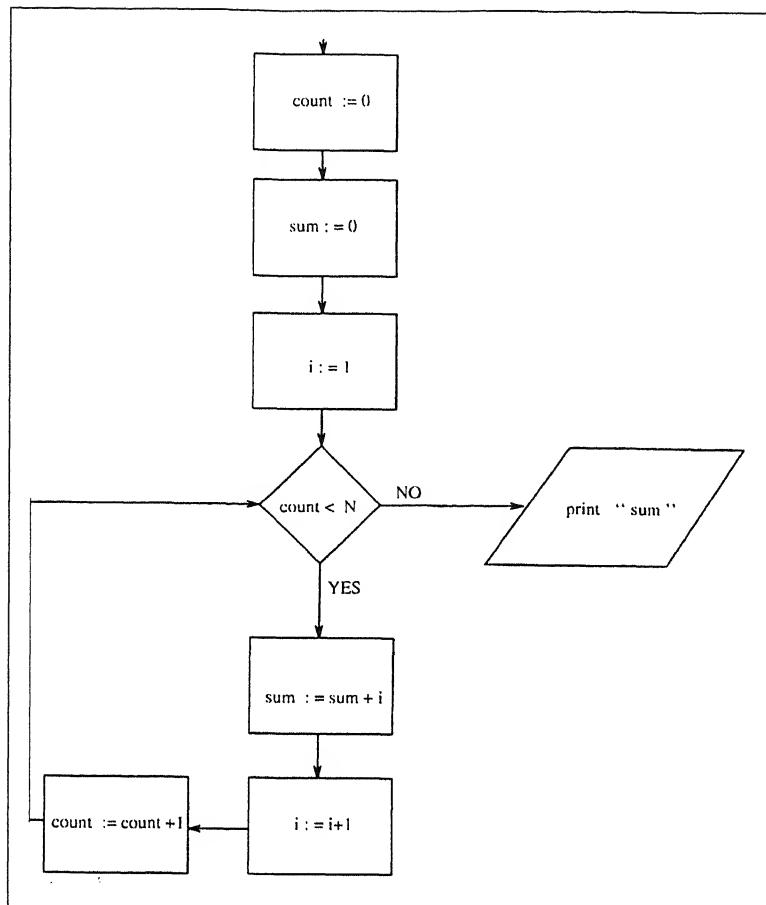
Figure 2 A flowchart to sum the first N natural numbers (N to be read separately).

Figure 3 Box 'read N' to be composed with the flowchart of Figure 2.
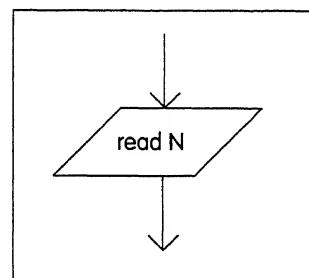
Table 1. Textual representation of the flowchart

count : = 0 ;

sum : = 0 ;

i : = 1 ;

while (count < N) *do*

      sum : = sum + i ; (* sum contains the sum of first *i* numbers *)

      i: = i + 1; (* increment *i* to get the next number *)

      count: = count + 1; (* count counts the numbers added *)

endwhile; (* sum contains the sum of first *N* numbers *)

print sum ;

● Let us suppose that n does not exactly divide m. Then, we have $m = p \times n + r$ for some p and $0 < r < n$. We can conclude that x must divide n and r. This follows since x must divide both the numbers m and n. Can we say anything stronger? Yes, we can say that $gcd\ (m,n)$ is the same as $gcd\ (n,r)$ (follows from the definition of gcd). The same argument can be applied for r and n. Note that the bound on the candidate for x gets reduced each time; now x is bounded by r. This is a crucial step in ensuring that the algorithm terminates.

Assuming m is greater than or equal to n, the flowchart for computing the gcd is shown in *Figure 4*. The operator rem used

Figure 4 Flowchart for computing gcd (m,n) using Euclid's algorithm.

The greatest common divisor (gcd) of two positive integers m and n is the largest positive number that divides both m and n.
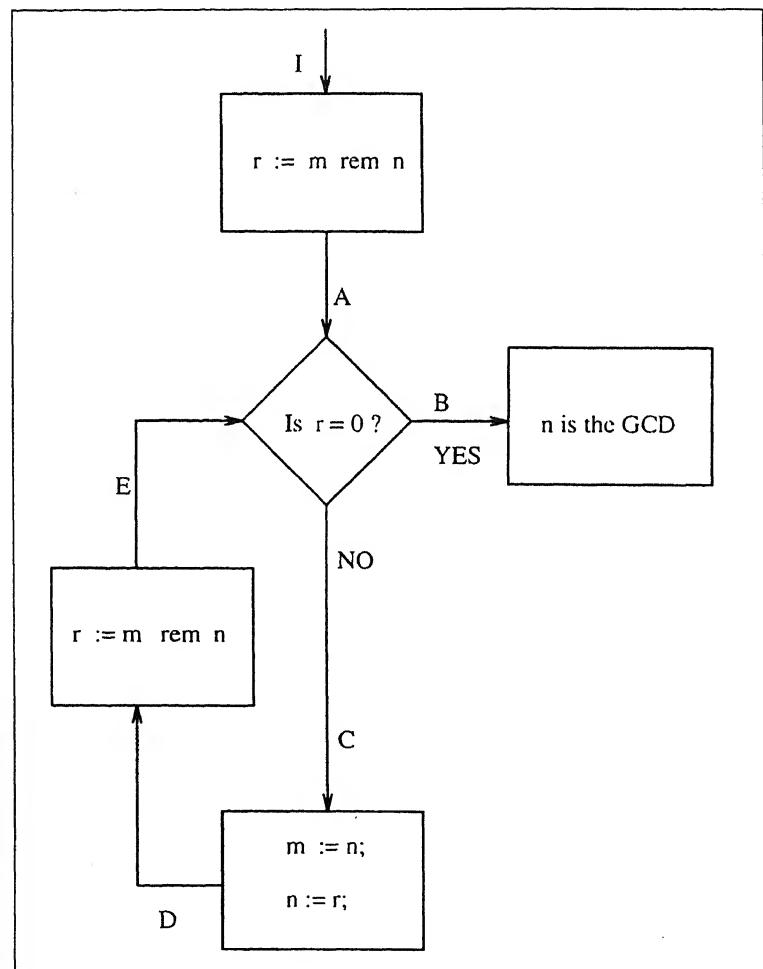
Table 2. Trace of gcd (8,6)	
Step	Action
1.	$r := 8 \text{ rem } 6 = 2$
2.	$r \neq 0 \longrightarrow m := 6; n := 2; r := 6 \text{ rem } 2 = 0$
3.	$r = 0 \longrightarrow$ '2' is the gcd

here is defined by: $p \text{ rem } q = $ remainder obtained on dividing p by q. The trace of Euclid's Algorithm for $m = 8$ and $n = 6$ is shown in *Table 2*.

Now let us see how we can informally argue that the algorithm indeed computes what we want. For convenience, we have labelled the arrows in the flowchart. By observing the flow of information, we can assert the following facts at the labels:

- At label A: r is set to the remainder obtained on dividing m by n. Hence, $0 \leq r < n$; m and n remain unchanged (i.e., $m = p \times n + r$ assuming $m \geq n$).
- At C: the remainder r is not equal to zero.
- At D: m is set to n and n is set to the remainder. Also, we have $m > n$. Can we say that the *gcd* of the original m and n and the new m and n are the same? From the discussion given above, we can indeed assert this statement.
- At B: The remainder r is equal to zero — leading to the *gcd*.

Example 3: Computing a factorial.

The familiar definition of *factorial* is

$$fact(n) = n! = 1 \times 2 \times \ldots \times n, \quad n \geq 0.$$

How do we derive an algorithm for computing *fact*? We first express *fact* (i) in terms of the *fact* (j) for $j < i$. Note that

fact (1)	$= 1$	(1)
fact (i)	$= 1 \times \ldots \times i$	(2)
fact ($i + 1$)	$= 1 \times \ldots \times i \times (i+1)$	(3)

The algorithm for computing the factorial (fact) uses the recurrence relation that fact $(i + 1)$ equals fact $(i) \times (i+1)$.

Table 3. Algorithm for Computing a Factorial

```
fact := 1;
i := 0;
while i ≤ N do
    i := i+1;
    fact := fact * i
end while
print fact;
```

Assuming *fact* (0) = 1, and combining (2) and (3) we get the following relations (recurrence):

$$fact\,(0) \quad = \quad 1 \tag{4}$$
$$fact\,(i+1) \quad = \quad fact\,(i) \times (i+1) \tag{5}$$

The reader may observe that various interesting programs can be developed along the same lines for computing the 'sine' function from its series, the 'Fibonacci numbers', etc.

Now, we can get a simple algorithm using the relations (4) and (5). In the algorithm, we start with an initialization of *fact* (0) to be 1. The successive factorials can then be obtained by multiplying the immediate preceding factorial (computed in the previous step) by the next natural number. The algorithm is described in *Table 3*.

Example 4: Finding the 'integer' square root.

We devise an algorithm to find the approximate (integer) square root of a fixed number $n \geq 0$. For example, the integer square root of 4 is 2, integer square root of 5 is 2, and integer square root of 10 is 3. That is we have to find an a such that

$$a^2 \leq n < (a+1)^2 \tag{6}$$

The basic scheme would be to start from a good guess and move on to the next guess if the number chosen does not satisfy the required property. It is important that when we move from the current guess to the next guess we do not miss the actual number we are looking for. Thus, starting with 0 as the first guess and incrementing it by one every time till (6) is satisfied, will eventually yield the result. But it will be too 'expensive'. We can learn

---

Table 4. Finding the Integer Square Root.

(* Finding the integer square root of n *)

$a := 0$;	(* lowest guess *)
$b := n + 1$;	(* largest guess *)
while $(a+1) \neq b$ do	(* get the average guess *)
$d := (a + b) \div 2$;	(* $\div$ denotes integer division*)
if $(d * d) \leq n$ then	
$a := d$	(* refined lower guess *)
else $b := d$	(* refined largest guess *)
endif	
endwhile	

---

something from the relation (6) itself. We simultaneously guess a *lower* bound (say l) and an *upper* bound (say u) and update these two bounds appropriately. At the initial stage, 0 is a candidate for l and $n+1$ is a candidate for u. Next, how do we update l and u? By taking the square root of the numbers involved in the relation (6) we can derive the following relation

$$a \leq \sqrt{n} < (a+1) \tag{7}$$

Thus, a is bounded above by $\sqrt{n}$. Let us try to reduce the interval (l, u) by half, by setting l or u to $(l + u)/2$ such that the condition $l < u$ is still satisfied. The reader can check that this strategy will not skip the number we are looking for. Note that l will never reach the upper bound. This idea has been used to develop the algorithm described in *Table 4*.

Example 5: Finding an 'approximate' square root.

In the previous section, we developed an algorithm for finding the integer square root of a number. The integer square root can be considered as a crude approximation to the square root of a number. Let us see whether we can modify the above technique and compute the square root of any positive number such that it differs from the actual square root by at most some given *tolerance*

To find the integer square root the basic scheme would be to start from a good guess and move on to the next guess if the number chosen does not satisfy the required property. It is important that when we move from the current guess to the next guess we do not miss the actual number we are looking for.

limit. Since square roots of natural numbers need not be natural numbers, such a modification will permit us to find the square root of decimal numbers also. It may be pointed out that in general, we cannot compute the exact value of the square root of a number, as the number may not be representable in the given machine accuracy.

Now, we will adapt the above algorithm (given in *Table 4*) and compute the approximate square root of a number. Let us assume that $x_0 > 0$ is the first guess of the square root of the given number a; a is assumed to be a positive non-zero number. Then, a/x_0 is also an approximation to $\sqrt{a}$ and

$$a/x_0 < \sqrt{a} \quad \text{if } x_0 > \sqrt{a}$$
$$a/x_0 > \sqrt{a} \quad \text{if } x_0 < \sqrt{a}.$$

The interesting fact is that the average of x_0 and a/x_0, say x_1, is also an approximate square root and satisfies the property

$$x_1 \geq \sqrt{a}$$

The equality holds only if $x_0 = \sqrt{a}$. Note that it is not necessary that x_0 be greater than or equal to $\sqrt{a}$. We can repeat the process of obtaining the next approximate square root; then, the $(i+1)^{th}$ approximation (denoted by x_{i+1}) is given by

$$x_{i+1} = (x_i + a/x_i)/2$$

The fact that the new approximate square root is better than the earlier one follows from:

$$x_1 \geq x_2 \geq \ldots \geq x_{i+1} \geq \sqrt{a}$$

From this relation, we infer that the value gets refined through the process of finding the next approximation from the current one. The successive approximates of x_i are referred to as the *iterates.* Do note that each iterate is better than the earlier ones. The important question is: "When do we stop?" We stop when the iterates stop decreasing; in the above case, they stop decreasing when there are no more representable values between $\sqrt{a}$ and x_i with the given machine accuracy. Suppose *error* represents the accuracy to which

Table 5. Finding the Approximate Square Root

```
(* Finding the approximate square root of A *)
a := A ;                           (* A is the given number *)
E := error,                        (* error is the given accuracy *)
xold := X₀ ;                       (* initial guess *)
xnew := (xold + a / xold) / 2 ;    (* refined root *)
while (xnew - xold) > E do
        xold := xnew,
        xnew := (xold + a / xold) / 2 ;  (* refined root *)
endwhile
```

the number can be represented in the given computer. Then, we can stop whenever $(x_{i+1} - x_i)$ is less than or equal to this quantity. Assuming that we have been given an *initial guess* and an *error* which can be tolerated, the program for finding the approximate root is given in *Table 5*.

Note the following:

- The division operator '/' used in *Table 5* denotes the usual division operation and not the integer division operation used in *Table 4*.
- Unless the initial guess is the correct guess, the equality in the relation among the iterates does not hold. Thus, if we start with an incorrect guess even for a natural number having an exact square root, we will not get the exact root using this method.
- The number of iterations before the program terminates depends on the starting values (initial guesses); it is of interest to note that there are procedures to arrive at these initial guesses for the technique discussed above.

The method described above for computing the approximate square root is referred to as Newton's method for finding $\sqrt{a}$ after the famous English mathematician Isaac Newton.

In *Table 5*, we have essentially solved the nonlinear equation

$x^2 = a$. The method can be extended to find the nth root of the equation $x^n = m$ and it is usually referred to as the Newton-Raphson method.

Discussion

In the previous sections and the previous article, we have learnt several constructs such as: assignment and basic commands, sequential composition, *if-then-else*, and the *while* construct. We can categorize these constructs into two classes:

- *Imperative Commands*: These are instructions to the processor. Constructs such as assignment and other basic commands belong to this class.
- *Control Commands*: These are commands which reflect the way in which the various instructions are sequenced. The *if-then-else* statement provides conditional sequencing of instructions, and the *while*-construct provides conditional sequencing based repeatedly on a given condition. These constructs are referred to as *control structures. The* control structures abstract the way the commands can be executed on a machine. Such an abstraction is often referred to as *control abstraction*. It must be evident to the reader that one can devise various other control structures. For instance, one can devise a construct where the control enters the statement block first and is tested at the end of the statement block execution. This is different from the while-construct where a condition is tested before entering a statement block. One such construct is the *repeat-until* construct. For example, *repeat* S *until* B *endrepeat* can be interpreted as: Repeatedly execute S until the condition B holds. Thus, when the statement terminates, we can conclude that B holds.

Having looked at the above basic constructs, it is natural to ask the following questions:

- Are the above mentioned constructs general for all programming purposes and if yes, in what sense?

Address for correspondence
R K Shyamasundar
Computer Science Group,
Tata Institute of
Fundamental Research,
Homi Bhabha Road,
Bombay 400 005, India.

- In the description of algorithms and programming languages, what is the role of control abstraction?
- What are the inherent limitations of the algorithmic processes?

In future articles in this series, we will show that these constructs are powerful and can be used to encode any algorithm. In the next article, we will discuss procedural abstraction and one very widely used programming technique called recursion in the context of procedural abstraction. We will also provide a relative comparison of the iterative and the recursive constructs in the description of algorithms.

We can categorize the constructs studied so far into two classes: *imperative commands,* which are instructions to the processor and *control commands,* which reflect the way various instructions are sequenced.

Suggested Reading

D E Knuth. Art of Computer Programming. Volume 1. Addison-Wesley Publishing Co. 1972.

E W Dijkstra. A Short Introduction to the Art of Programming. Computer Society of India. 1977.

G Polya. How to Solve It. Princeton University Press. 1973.

R G Dromey. How to Solve it by Computer. Prentice-Hall of India, New Delhi. 1990.

The ultimate folly... "The worst thing that can happen — will happen (in the 1980s) — is not energy depletion, economic collapse, limited nuclear war, or conquest by a totalitarian government. As terrible as these catastrophes would be for us, they can be repaired within a few generations. The one process ongoing in the 1980s that will take millions of years to correct is the loss of genetic and species diversity by the destruction of natural habitats. This is the folly our descendants are least likely to forgive us.". *(E O Wilson, Harvard Magazine, January-February 1980).*

The discovery of the Mobius strip ... In 1858, a scientific society in Paris offered a prize for the best essay on a mathematical subject. In the course of coming up with an essay for this competition, August Ferdinand Mobius, a mathematician in Leipzig, Germany, 'discovered' the surface that now bears his name. It is called the Mobius-strip.

Geometry

3. Towards a Geometry of Space and Time

Kapil H Paranjape

After spending about a decade at the School of Mathematics, TIFR, Bombay, Kapil H Paranjape is currently with Indian Statistical Institute, Bangalore.

In the first two articles of this series the author described Euclidean geometry, coordinate geometry, trigonometry and measure theory. In this article he introduces non-Euclidean geometry and discusses tangents to curves and surfaces. These seemingly different notions will be brought together in the future when he discusses differential geometry.

Working at Parallel Purposes

Parallel: A pair of lines in a plane is said to be parallel if they do not meet.

Mathematicians were at war with one another because Euclid's axioms for geometry were not entirely acceptable to all. Archimedes, Pasch and others introduced further axioms as they thought that Euclid had missed a few, while other mathematicians were bothered by the non-elementary nature of the parallel axiom. They wondered if it could be proved on the basis of the other axioms.

One form of the Parallel Axiom of Euclidean Geometry is:

There is exactly one *line that is parallel to a given line and passes through a given point not on it.*

¹Captured by the following verse or worse:
*If the theorem isn't true
Then the sky isn't blue
Which is so absurd
That the truth you must have heard*

By constructing a pair of right angles it is not hard to show (using the remaining axioms) that there is at least *one* parallel line as required by the above axiom (exercise). So we are tantalisingly close and only need to show that the line is unique. Many mathematicians devoted much of their careers to this problem. By the method of *reductio ad absurdum* (reduction to absurdity)[1] they
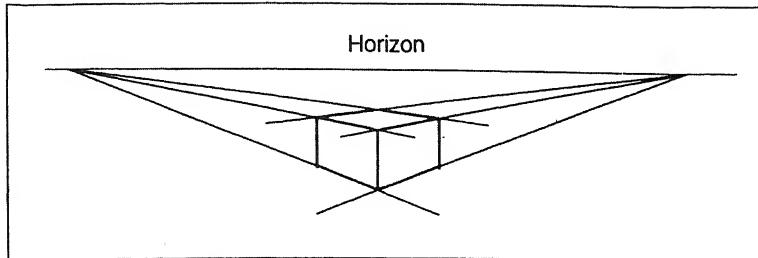
attempted to derive (from the supposedly wrong hypothesis that there is another parallel) a number of 'results' that would seem absurd. The most successful in this was Saccheri. Many of his 'results' actually became theorems in non-Euclidean geometry — results which he thought were wrong!

Indeed, Lobachevsky and Bolyai showed that there is a perfectly valid geometry in which the parallel axiom is replaced by the following:

There are at least two lines that are parallel to a given line and pass through a given point not on it.

Subsequently, Poincaré, Klein, Beltrami and others refined non-Euclidean geometry. It was shown that Euclidean and non-Euclidean geometry are equi-consistent — one is as consistent as the other.

From a different perspective artists had all along pointed out that parallel lines *do meet* at the horizon *(Figure 1)*. In fact all pairs of coplanar lines meet and parallel lines are singled out by the fact that their point of meeting is at the horizon. The horizon is itself another line (called the line 'at infinity'). A perfectly consistent geometry can thus be formed with the axiom:

There are no parallel lines. Any pair of coplanar lines meet.

Of course, the axioms of separation need to be modified somewhat to accommodate the fact that every line becomes 'circular'. This geometry is the *projective geometry* of Poncelet which reappeared as the *elliptic geometry* of Riemann.

Euclidean and non-Euclidean geometry are equi-consistent — one is as consistent as the other.
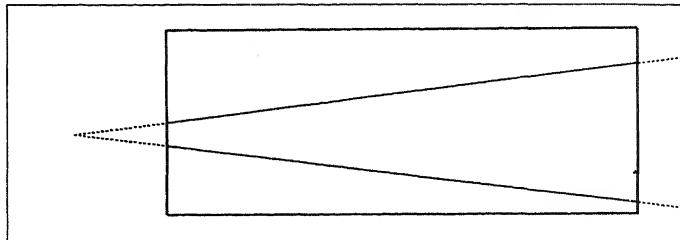
The Local Axioms

Incidence. Each pair of distinct points determines a unique line and so on.

Dimension Any two planes that meet have at least two points in common. There are four non-coplanar points.

Separation . Each point on a line divides the line into two rays; each line on a plane divides the plane into two half planes and so on.

LUB property . Given a sequence of points A_n and a point B so that A_{n+1} is between A_n and B, i.e. B is an upper bound for the A_n's, there is a point C which is a least upper bound.

The existence of so many parallel choices for the parallel axiom appears to spell trouble since it calls into question the introduction of co-ordinate geometry. (We recall from the first article of this series that the introduction of co-ordinates depended on the parallel axiom.) However, it was shown by Klein, Beltrami and others that coordinates are a natural consequence of the axioms of 'local space'. (Note that there can be no talk of parallels since even lines which might meet 'far away' do not meet in the given region: see *Figure 2*). More precisely we restrict ourselves to the axioms of incidence, dimension, separation and the least upper bound property (LUB) — axioms that appear to be satisfied by a small region of space surrounding us (see box). It turns out that there is a natural way to embed such a geometry in co-ordinate geometry (so that the lines embed as lines, planes as planes and so on). Thus we can safely return to the study of co-ordinate geometry secure with the knowledge that all non-Euclidean phenomena can be found there.

A question that perhaps still nags practical-minded people is whether the geometry of the space around us is Euclidean or not. The first person to try to test this was Gauss but the scale he chose was not large enough to show the non-Euclidean nature of space. In this century as a test of Einstein's theory of gravitation the 'curved' nature of space was finally shown.

Time to Take Off on a Tangent

Tangent line. A line with the *maximal order of contact* with the given curve at the given point among all lines through this point.

Since time immemorial — or at least since time became measur-

able, people have wanted to know how fast or slow things change or move. A crude way of doing this is to measure the change that takes place in a specified amount of time or to use a stop watch to measure how long it takes to achieve a specified amount of displacement. In the interests of accuracy one must use smaller and smaller time periods. To follow this logic to its limit one must use a zero time period but that is apparently absurd. The word *limit* in the previous sentence gives a clue to the correct approach — and this is how 'instantaneous velocity' or derivative was finally given analytical meaning. However, this analytical definition of derivative in terms of limits really came much later with the work of Cauchy and Weierstrass. Newton and Liebnitz argued on the basis of other concepts.

One was the use of 'infinitesimals' — or infinitely small entities. This was vehemently argued against at the time[2] but is in fact quite a straight-forward algebraic method which can be made perfectly valid when the position is an algebraic (or more generally, analytic) function of time. As an example, to find the instantaneous velocity of an object occupying the position $(t+t^2, t^3)$ at time t we substitute $(t+h)$ in place of t,

$$((t+h) + (t+h)^2, (t+h)^3) = ((t+t^2)+(1+2t)h, (t^3) + (3t^2)h)$$

when h is an 'infinitesimal' such that h^2 is zero but h is itself not zero[3]. Now the coefficient of h, that is $(1+2t, 3t^2)$ gives the instantaneous velocity 'vector'.

The second method was based on what is now called the Liebnitz rule of derivation. There are in fact three rules

$$D(f+g) = Df + Dg \; ; D(\alpha f) = \alpha \, Df \; ; D(fg) = (Df)g + f(Dg)$$

where f and g are functions and α a constant. Combined with the 'choice' $Dt = 1$ we easily see that $D(t + t^2, t^3) = (1+2t, 3t^2)$. This is another algebraic method and one that has been used very fruitfully in commutative algebra in the recent past.

Newton was principally a mathematician whatever some other

Since time immemorial — or at least since time before measurable people have wanted to know how fast or slow things change or move.

[2] It is not accepted by many teachers today; so don't use this method in your exam papers!

[3] Those who find this difficult to digest can think of the matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ which has the same property.
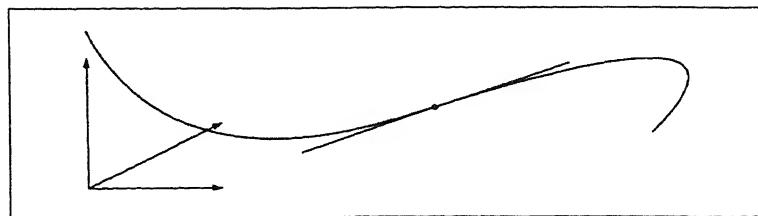
people may tell you! At least he was a geometer since he gave the geometric definition of instantaneous velocity by means of the tangent. First, one considers the trajectory as a curve with time plotted as one of the axes; in the example above we have the parametric curve $(x(t), y(t), t) = (t+t^2, t^3, t)$. The slope of the tangent line (as defined above) with respect to the time axis gives the derivative (*Figure 3*).

Thus the study of nebulous physical quantities like time (is it justified to treat it as one of the axes — i.e. a real variable?) and dubious analytic constructs like limits (at least until the work of Cauchy and Weierstrass) is replaced by the clear[4] geometrical notion of tangents.

[4] Clarity is clearly in the eye of the beholder!

Another way to look at tangents is by means of the 'apparent locus'. This is how the curve will appear at a specified point, to a person constrained to move along the given curve.

[5] As opposed to the usual notion of bird's eye view the apparent locus can perhaps be called *the ant's eye view*!
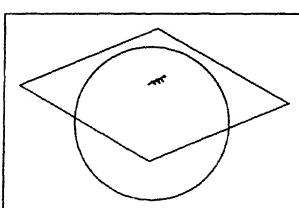
This becomes clearer if we consider higher dimensional tangents. Since an ant is constrained to move along the surface of the earth, which we may assume to be a sphere, at any position it imagines the earth to be planar — this plane being the tangent plane to the sphere at that point[5] (*Figure 4*). The locus of all tangent lines then makes up the 'apparent horizon'; this is how the horizon appears to an ant constrained to move along the given locus. These apparently simple notions play an important role in geometry since they allow us to study tangency even for more complicated situations. (Exercise for the adventurous: What is the apparent horizon of the locus $x^2+yz-x^3=0$ at the origin?)

Figure 4 Ant's eye view.

Differential calculus thus becomes the most important tool with

which one can study the more general loci that co-ordinate geometry allows us to introduce.

Summary

A non-Euclidean geometry is one where the notion of parallel line is changed to include a multiplicity of parallels. Even such a geometry can be given co-ordinates. Thus co-ordinate geometry wins the day. Differential calculus is the most useful tool in the study of co-ordinate geometry. This explains why most undergraduate studies in mathematics begin with *calculus and analytic geometry*.

We will now take a break for rumination (or to chew gum). When we return we shall see how Gauss and Riemann put together the above tools so that today even an ant can decide whether space is curved.

Suggested reading

H S M Coxeter. Non-Euclidean Geometry. University of Toronto Press. 1961.
  A good introduction to projective geometry and other non-Euclidean geometries.
D Hilbert. Foundations of Geometry. Open Court Publishers, La Salle, Illinois, USA. 1971.
  A more advanced treatment can be found in this book.
D Hilbert and S Cohn-Vossen. Geometry and the Imagination. Chelsea, NY, USA. 1952.
  A difficult but juicy book.
R Courant and H Robbins. What is Mathematics? Oxford University Press. 1941.
  A must-read book for a look at real mathematics.

Address for correspondence
Kapil H Paranjape,
Indian Statistical Institute,
8th Mile, Mysore Road,
Bangalore 560 059, India.

This series on Geometry will resume in May 1996.

The manifold genius of Helmholtz... The *Dictionary of Scientific Biography* lists Hermann Helmholtz's eminence in the following fields: energetics, physiological acoustics, physiological optics, epistemology, hydrodynamics and electrodynamics — without mentioning his work in non-Euclidean geometry.

Know Your Chromosomes

2. The Strong Holds of Family Trees

Vani Brahmachari

We continue the series with a discussion on family trees or pedigree analysis and the mapping of genes through linkage studies.

Kings, canine breeders and geneticists worry about pedigree. A cytogeneticist while dealing with a suspected genetic disorder asks the patient for a family history and draws a *pedigree chart* (*Figure 1*) which shows if there are other family members having the same disorder and whether there is any sex bias in its occurrence, i.e. does the disorder occur in males more often than in females or vice versa.

Chromosome Sets and Gene Locations

We know that most organisms including humans, are diploid which means that each gene is present in at least two copies. In fact, some genes in the human genome are present in multiple copies. You can well imagine that when one copy gets bad the other one can take over just as two musicians can fill in for each other in a concert. If one stops singing, the concert does not have to stop but can continue, perhaps as well as before. A similar situation occurs in the case of most genes on chromosomes 1 to 22, which are the autosomes. But the X and Y chromosomes in men are like single musicians singing in a concert; if this musician stops singing, the concert comes to an end. The X chromosome in the female still enjoys the luxury of having a second copy but males have only one X chromosome. They are haploid for X and Y chromosomes. Now does it make sense to say that any defect in a gene located on the X chromosome is expressed in males more often than in females? For most genes located on the autosomes we see no such sex bias.

Vani Brahmachari
is at the Developmental
Biology and Genetics
Laboratory at Indian
Institute of Science.
She is interested in
understanding factors
other than DNA sequence
per se, that seem to
influence genetic
inheritance. She utilizes
human genetic disorders
and genetically weird
insect systems to
understand this
phenomenon.

Terms italicized in the main text are explained in the glossary on page 39.
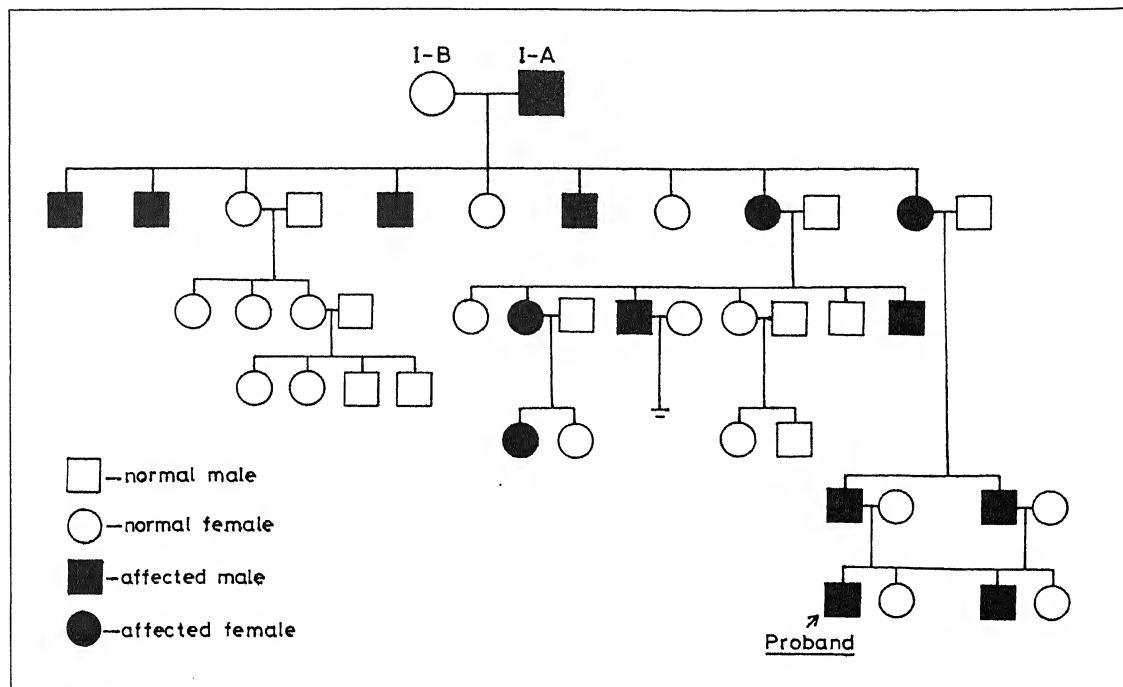
There are about 357 loci mapped on chromosome 1 so far (1st September, 1995). The word loci (singular *locus*) is used instead of genes because there can be a disease state mapped on a given chromosome, for which the specific gene has not yet been identified. *Locus* is the term for a chromosomal region where a trait is mapped. There may be more than one gene in the region responsible for the trait. For instance, there is a locus termed neuroblastoma on the short arm of chromosome 1 in band 32. The gene position is denoted as 1p32. The information known in this case is that a defect in this region results in an increased susceptibility to neuroblastoma. This band itself comprises about 3 million base pairs of DNA capable of housing 50-100 genes. Therefore the trait leading to neuroblastoma can be due to a defect in one or more genes in the region. On the other hand, 'Maple syrup urine disease-type II' can be called a gene because it is known that the gene codes for a polypeptide chain of a multiple subunit enzyme called alpha-keto acid dehydrogenase. In this particular example the gene sequence is also known. But often the terms locus and

There are about 357 loci mapped on chromosome 1 so far.

gene are used as synonyms.

Pedigrees Help Determine the Mode of Inheritance

Chromosome 1 is one in which the largest number of loci have been mapped. It is also the first human autosome on which a gene locus was mapped. As mentioned in the previous article (*Resonance*, Vol.1, No.1) there are several approaches to mapping genes. In this case I will use the example of antithrombin III (AT3) and describe gene mapping by the classical genetic approach.

Antithrombin is an α-globulin protein found in the *plasma*. It acts as the principal inhibitor of thrombin and other coagulation factors in blood. Deficiency of AT3 leads to disorders related to inappropriate blood clotting like 'thrombophlebitis' and 'acute aortic thrombosis'. Patients die prematurely because of blood clotting in vital arteries and veins. There are several families where this deficiency occurs. One such pedigree or family chart is shown in *Figure 1*. When the male marked as '*proband*' came to the doctor's attention, the family history of the patient and his parents as well as the AT3 phenotype in all available family members was documented. An analysis of the data presented in *Figure 1* reveals the following features:

> Deficiency of AT3 leads to disorders related to inappropriate blood clotting like 'thrombophlebitis' and 'acute aortic thrombosis'.

a) Both males and females in the family are affected.
b) The father I-A, but not all his children has the disease. Only about 50% of the children of an affected parent have the disorder.
c) When a child is affected, we see that at least one of the parents has the disease.
d) When both the parents are normal, none of the children are affected.

Based on these observations one can draw certain conclusions. This disorder is not sex chromosome linked because both males and females are affected. Only one of the two AT3 genes that the father I-A has is defective, as there are normal children among his progeny. Genetically he is said to be *heterozygous* for AT3 as against his wife who is *homozygous* for normal AT3. The observa-

tions (c) and (d) would lead one to conclude that it is a *dominant* disorder. This means that whenever an individual has one defective copy of the gene and the other is normal, he or she expresses the disorder. In technical terms the individual shows the disease *phenotype*. Thus, we conclude that AT3 deficiency is an autosomal dominant disorder.

Mapping by Linkage

Having determined the mode of inheritance, the next task was to localize AT3 to a specific chromosome. In experimental animals one can interbreed individuals with selected phenotypes. For instance, one can set up a mating between a brown eyed female with AT3 deficiency and black eyed male with normal AT3. Similar matings using several pairs of animals with different combinations of eye colour and AT3 gene can be set up. Let us assume that with respect to the phenotype of the AT3 gene (normal versus defective) and eye colour (brown versus black), the offspring always receives the parental combination of the two phenotypes; i.e. whichever offspring inherits AT3 deficiency also inherits brown eyes. In the example described here one concludes that AT3 and eye colour genes are close together and are not separated by *genetic recombination*. This is an instance of *genetic linkage* which means that AT3 is closely linked to the eye colour gene. Assuming that one knows the chromosomal location of brown eyes one could then map AT3 to the same chromosome. Of course, this approach is not possible in humans who marry with total disregard to the needs of geneticists!! In spite of this, the large size of human populations and the fact that individuals who show deviations from normal development and physiology are brought to medical attention, make it possible to obtain information to map human chromosomes. To identify the exact position of the gene on the chromosome other approaches are required, and these will be discussed in subsequent articles.

In the specific case of AT3, linkage studies were done with several known traits or markers in various families that had been kept on

The large size of human populations and the fact that individuals who show deviations from normal development and physiology are brought to medical attention, make it possible to obtain information to map human chromosomes.

Genetic Recombination

Recombination is the process of exchange of genetic material between *homologous chromosomes*. Recombination during meiosis results in the formation of gametes (sperms and eggs) with combinations of genes different from those present on parental chromosomes. The progeny of such matings will generate individuals with combinations of traits not present in the parents. The frequency with which recombinant individuals occur in a pedigree, provides an estimate of the distance between the genes. The greater the distance between two genes, the higher their recombination frequency; 50% being maximum, in which case the genes will behave as if they were on two different chromosomes. To map a new gene based on recombination frequency one should have a large number of known genes on each chromosome so that this parameter of distance and *recombination frequency* can be reliably assessed.
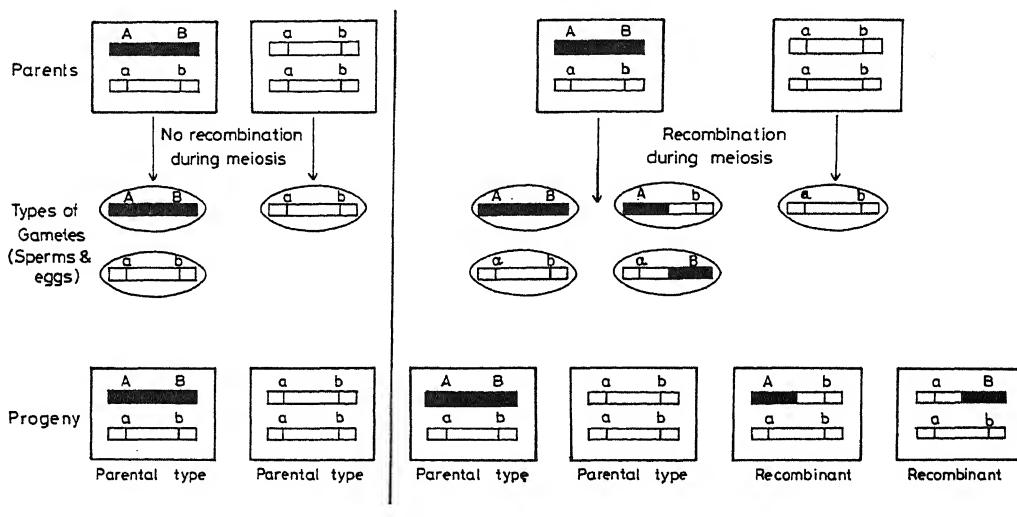
A and B are two traits that have two *alleles* (forms) A, a; B, b.
Each box represents an individual.

It was found that a blood group called 'Duffy' is inherited along with AT3 more often than expected by chance alone.

record with different investigators. It was found that a blood group called 'Duffy' is inherited along with AT3 more often than expected by chance alone. Duffy is an antigen on red blood cells like A, B or the Rh factor. It was originally detected in a patient named Duffy. The Duffy blood group represented as Fy had been mapped to chromosome 1. In this particular case it was seen that

in several pedigrees or families, Duffy blood group was present in those individuals who also had an abnormal looking chromosome 1. Localization of Duffy to chromosome 1 was based on this observation and the linkage of Duffy to another locus previously mapped to this chromosome. This was the only approach used to map genes on human chromosomes till modern methods of molecular biology were developed and applied. In the case of AT3 these methods have helped in localizing it to a defined band on chromosome 1 and in identifying the protein product encoded by the gene. The first description of the disorder was made around the year 1965, it was mapped to chromosome 1 in 1978 and the gene sequence was identified in 1983. One can imagine the effort required to complete these analyses which involve chromosomal mapping, identifying the gene, sequencing it and understanding its function! But methods of disease management, principally by the injection of anticoagulants, have not changed much over the decades. What good is it to dissect the disease process to its ultimate molecular basis when one cannot correct it? It is hoped that a molecular understanding would lead us to the correction of gene defects at the DNA level. I will discuss some of these aspects in articles at the end of this series.

As I mentioned before there are more than 300 genes mapped on chromosome 1. There are source materials that provide all the information known about these genes. What I have done here is to provide an abbreviated map which indicates the genes known to cause specific disorders (*Figure 2*) and the biochemical pathways they affect (*Table 1*).

The challenging route from chromosomal location to identifying the gene and pinning down the functional defect is too intricate to describe in one article. Future articles will deal with this process, one aspect at a time. Just as hybrid plants and animals have been created for improving farm products, hybrid cells have served as one of the essential tools for genetic analysis. We will examine this approach in the next article.
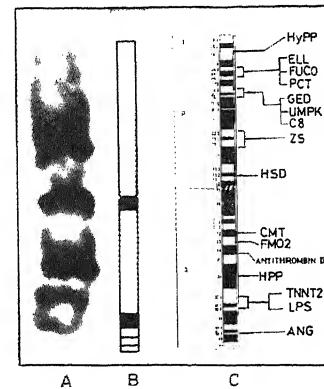
Figure 2 A diagrammatic representation of chromosome 1, as seen under a light microscope after giemsa banding (A), a diagrammatic representation indicating the bands used as landmarks for chromosome identification (B), schematic showing the position of genes listed in Table 1(C).

Table 1 A Listing of Representative Loci Mapped to Chromosome 1.

Gene/disorder	Chromosomal Location	*Mode of Inheritance*
1. Fucosidosis (FUCO)	1p 34	*Autosomal recessive*

The disorder is due to deficiency of fucosidase enzyme leading to accumulation of the sugar fucose in tissues. Neurological deterioration, growth retardation, seizures and early death are the manifestations of the disease.

2. UDP-galactose-4-epimerase deficiency (GED)	1p 32	*Autosomal recessive*

The enzyme epimerase is involved in the interconversion of galactose and glucose. Deficiency leads to seizures, requires external galactose.

3. Liver, Alkaline Phosphatase, Hypophosphatasia (HyPP)	1p 36	*Autosomal recessive*

Enzyme acts as a lipid anchored ectophosphatase. Seizures due to accumulation of pyridoxal-5'- phosphate which interferes with neurotransmitter levels. One of the forms is fatal in infancy. Adult form leads to skeletal abnormalities and premature shedding of teeth.

4. Porphyria cutanea tarda (PCT)	1p 34	*Autosomal dominant*

Disease results in light sensitive skin lesions, fragile skin due to deficiency of uroporphyrinogen dicarboxylase, an enzyme involved in biosynthesis of heme. A rare instance where an enzyme deficiency is dominant in its effect.

5. Complement component-8 C8 deficiency (C8)	1p 32	*Autosomal dominant*

Deficiency leads to defect in immune system resulting in frequent bacterial infections specially Neisseria and meningo coccus.

6. Uridine monophosphate Kinase (UMPK)	1p 32	*Autosomal recessive*

UMPK catalyses conversion of uridine monophosphate to uridine diphosphate. Deficiency leads to an impairment of effective immune reactions resulting in susceptibility to infections.

7. Peroxisomal membrane protein. PX MP1 Zellweger syndrome (ZS)	1p 22-p21	*Autosomal recessive*

Peroxisomes are single membrane bound organelles of cells, involved in oxidation of fatty acids. Deficiency of PXMP1 leads to absence of peroxisomes in liver and kidney and to early death.

Gene/disorder	Chromosomal Location	Mode of Inheritance
3-ß-hydroxy steroid dehydrogenase delta isomerase (HSD)	1p 13.1	*Autosomal recessive*
enzyme is involved in the biosynthesis of testosterone. Defect in the enzyme leads to excretion and nplete masculinization in males and mild virilism in females.		
Elliptocytosis (ELL)	1p 34.2-p33	*Autosomal recessive*
disorder is due to a defective membrane protein of red blood cells. Results in elliptic shape of RBCs, mia and gall stones. Removal of spleen relieves the symptoms.		
Charcot-Marie-Tooth Disease (CMT)	1q 22	*Autosomal dominant*
locus codes for a myelin protein. Mutation in the gene leads to neural disorders like CMT, and orineural polypathy.		
Flavin containing mono oxygenase 2 (FMO2)	1q 23-25	*Autosomal recessive*
enzyme mediates oxidation of amino-trimethylaminase (TMA) derived from diet. Defect leads to etion of TMA in urine, fishy odour of the body, anaemia and psychosocial problems.		
Hypokalemic periodic paralysis (HPP)	1q 31-32	*Autosomal dominant*
gene codes for a calcium channel protein that controls entry and exit of calcium ions. The defective gene s to weakness and paralytic attacks accompanied by low levels of potassium in blood.		
Cardiac Troponin-T2 (TNNT2)	1q 32	*Autosomal recessive*
onin is a muscle protein. This specific variety is made only in heart muscles. Defect in the gene manifests irdiomyopathy.		
Lip-pit syndrome (LPS)	1q 32	*Autosomal dominant*
pecific gene is identified. But a deletion or loss of DNA sequences from this region results in cleft palate deformities of the upper lip.		
Angiotensinogen (ANG)	1q 42-q43	*Autosomal recessive(?)*
gene codes for angiotensinogen which is a precursor of angiotensin II involved in elevating blood sure.		
bbreviations given in parentheses are used to indicate the gene loci in Figure 2.		

Suggested reading

Daniel L Hartl. Human Genetics. Harper and Row Publishers New York, Cambridge, London. 1983.

Morroe W. Strickberger. Genetics. MacMillan Publishing Company New York, Collier Mac Millan Publishers, London. 1976.

Walter Bodmer and Robin Mckie. The Book of Man. The Quest to Discover Our Genetic Heritage. Little Brown and Company, United Kingdom. 1994.

Victor A McKusick. Human Genetics. Prentice Hall Foundations of Modern Genetics Series. Prentice Hall of India Private Limited, New Delhi. 1968.

Victor A McKusick. Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive and X-linked Phenotypes. Vol I and II Tenth Edition. The Johns Hopkins University Press, Baltimore and London. 1992

Address for correspondence
Vani Brahmachari,
Developmental Biology and
Genetics Laboratory, Indian
Institute of Science,
Bangalore 560 012, India.

Glossary

Allele: One of an array of possible forms of a given gene, which can be distinguished by their differing effects on the manifestation of a genetic trait.

Dominant and Recessive allele: An *allele* that expresses its effects even in the presence of a different form of the same gene is called a dominant allele. The allele whose effects are overridden by the dominant allele is said to be recessive.

Genotype: The specific allelic constitution of genes of an organism or a cell.

Genetic linkage: This means that two or more genes in question are located on the same chromosome. The proximity of the genes is estimated by recombination frequency.

Heterozygosity: A condition of having one or more pairs of dissimilar alleles.

Homozygosity: This means having identical rather than different alleles in the corresponding loci of homologous chromosomes.

Homologous chromosomes: Chromosomes that pair during meiosis. Homologous chromosomes contain the same linear sequence of genes and as a consequence, each gene is present in at least two copies.

Locus: (plural loci) The position that a gene occupies on a chromosome.

Pedigree analysis: Analysis of ancestral history or genealogical register or a family tree drawn to show the inheritance patterns for specific phenotypic characters.

Phenotype: The detectable manifestations of a *genotype* in conjunction with the environment.

Plasma: Fluid portion of the blood made up of proteins.

Proband: The person who is brought to the attention of the clinician for a potential genetic disorder. Starting from him/her the family tree is traced.

Recombination: The process that generates progeny with combination of alleles on a given chromosome, different from those that occur in the parents.

Recombination frequency: The proportion or percentage of products of recombination. This frequency is used as a guide in assessing the relative distances between loci on a genetic map. he proportion or percentage of products of recombination. This frequency is used as a guide in assessing the relative distances between loci on a genetic map.

From Matter to Life: Chemistry?!

Jean-Marie Lehn

In the beginning was the Big Bang, and physics reigned. Then chemistry came along at milder temperatures; particles formed atoms; these united to give more and more complex molecules, which in turn associated into organized aggregates and membranes, defining primitive cells out of which life emerged.

Chemistry is the science of matter and of its transformations, and *life* is its highest expression. It provides structures endowed with properties and develops processes for the synthesis of structures. It plays a primordial role in our understanding of material phenomena, in our capability to act upon them, to modify them, to control them and to invent new expressions of them.

Chemistry is also a science of transfers, a communication centre and a relay between the simple and the complex, between the laws of physics and the rules of life, between the basic and the applied. If it is thus defined in its interdisciplinary relationships, it is also defined in itself, by its object and its method.

In its method, chemistry is a science of interactions, of transformations and of models. In its object, the molecule and the material, chemistry expresses its creativity. Chemical synthesis has the power to produce new molecules and new materials with new properties. New indeed, because they did not exist before being created by the recomposition of atomic arrangements into novel and infinitely varied combinations and structures.

The first chemical manipulations were those profound and complex thermal transformations by which 'the raw' becomes 'the cooked'. Thus, before being understood, chemistry was practised in the cooking of food; in the fermentation of dough and of drinks; in the metallurgy of trinkets, of tools and weapons; in the extraction of natural substances, perfumes, colours, philtres and drugs.

Jean-Marie Lehn worked with Guy Ourisson for his doctorate degree. Later he was a post-doctoral fellow with Robert B Woodward at Harvard University. He had the distinction of being one of the coworkers who participated in the total synthesis of vitamin B_{12}. Lehn then joined the University Louis Pasteur at Strasbourg. Since his appointment at the College de France in Paris he is actively pursuing research in both Institutes. He is recognised as one of the pioneers in the development of *Supramolecular Chemistry;* (he coined the phrase). His research work encompasses a wide range of chemistry. Jean-Marie Lehn has received numerous awards and honours, including the 1987 Nobel prize in chemistry.

Chemistry was for a long time empirical and descriptive. In the days of alchemy, all matter resulted from the admixture of varying proportions of the four basic elements: fire, air, earth and water. Through more and more refined analysis, chemistry became the composition of substances, then the combination and linkage of atoms. When empirical formulae gave way to exact formulae, the problem of the geometry of arrangement of atoms in molecules arose, whence emerged the notion of structure. With van't Hoff and Le Bel, the planar chemical representations unfolded and left the plane. The molecule became architecture, and the elaboration of molecular structures became mastery of space. With the work of Pasteur, optical activity found its explanation in the asymmetry of structures; thus arose the body of molecular chirality.

The alchemists' dream of transmutation became reality in *chemical synthesis*. At first an empirical knowhow, this science of molecular manipulation established its power with Friedrich Wöhler's synthesis of urea in 1828: "Über Kunstliche Herstellung des Harnstoffes", (On the artificial preparation of urea), by which it was proved that natural substances were chemical compounds like any other, accessible in the laboratory without the intervention of a living organism. On 28 February 1828, Wöhler wrote to Berzelius "I can make urea with no need of a kidney, or let alone an animal, be it a man or a dog".

Organic synthesis grew rapidly, continually adding to its panoply of tools with each new reaction discovered, allowing access to innumerable new compounds, along with the refinement of strategies for obtaining more and more complex natural substances in the laboratory. A whole series of brilliant achievements, where elegance of strategy combined with feats of efficiency and selectivity, led to the great syntheses of the last 50 years — notably, to what is considered to be the epitome, the synthesis of vitamin B_{12}, due to the combined efforts of Robert Burns Woodward and Albert Eschenmoser, assisted by a hundred or so collaborators. We are a long way from Wöhler's urea! On the one hand there is a planar molecule with 4 atoms (not counting hydrogens), and on the other,

Wöhler's synthesis of urea in 1828 proved that natural substances were chemical compounds like any other, accessible in the laboratory without the intervention of a living organism.

a set of 93 atoms and extremely difficult stereochemical problems; between the two lies a century and a half (see *Figure 1*).

Molecular chemistry, thus, has established its power over the covalent bond. The time has come to do the same for non-covalent intermolecular forces. Beyond molecular chemistry based on the covalent bond there lies the field of *supramolecular chemistry*, whose goal it is to gain control over the intermolecular bond.

It is concerned with the next step in increasing complexity beyond the molecule towards the supermolecule and organized polymolecular systems, held together by non-covalent interactions.

It is a sort of molecular sociology! Non-covalent interactions define the intercomponent bond, the action and reaction, in brief, the behaviour of the molecular individuals and populations: their social structure as an ensemble of individuals having its own

Elegance of strategy combined with feats of efficiency and selectivity, led to the synthesis of vitamin B$_{12}$, due to the combined efforts of Robert Burns Woodward and Albert Eschenmoser, assisted by a hundred or so collaborators.

Figure 1 *Two milestones in organic synthesis: Urea (left) and Vitamin B$_{12}$ (right).*
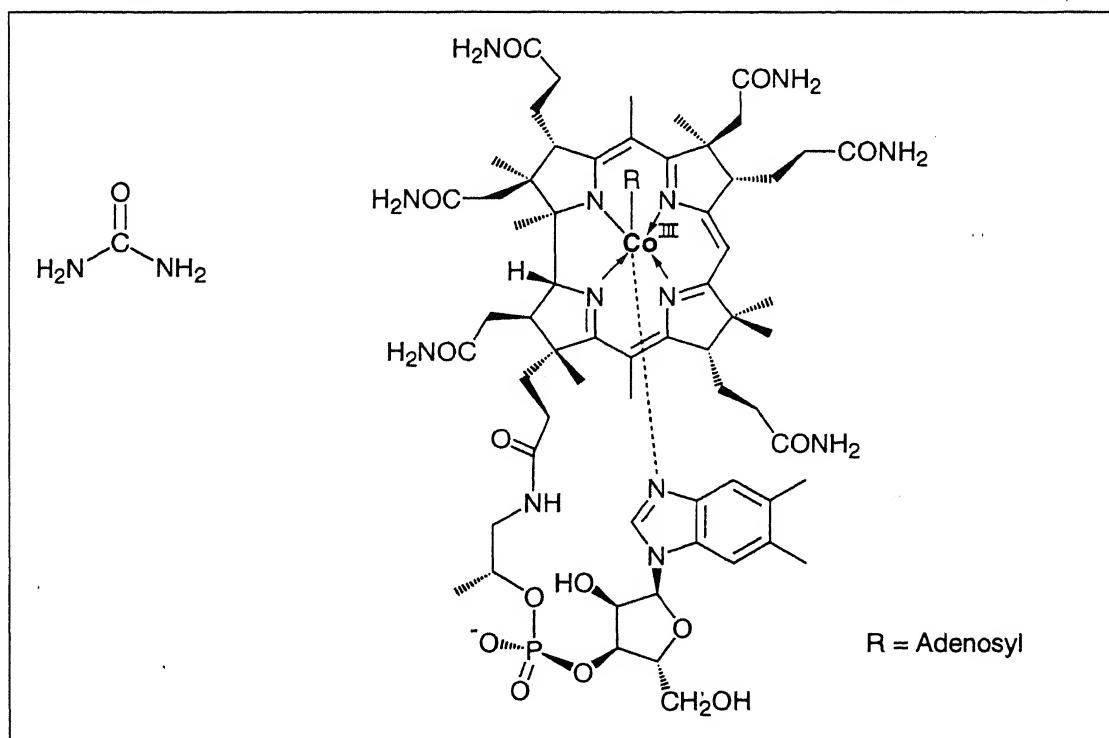
R = Adenosyl

organization; their stability and their fragility; their tendency to associate or to isolate themselves; their selectivity, their "elective affinities" and class structure, their ability to recognize each other; their dynamics, fluidity or rigidity of arrangements and of castes, tensions, motions and reorientations; their mutual action and their transformations by each other.

When a substrate binds to an enzyme or a drug to its target, when signals propagate between cells, highly selective interactions occur between the partners and control the process. Supramolecular chemistry is concerned with the study of the basic features of these interactions and with their implementation in specially designed non-natural systems.

Molecular interactions form the basis of the highly specific recognition, reaction, transport, regulation, etc; processes that occur in biology, such as substrate binding to a receptor protein, enzymatic reactions, assembling of multiprotein complexes, immunological antigen-antibody association, intermolecular reading, translation and transcription of the genetic code, regulation of gene expression by DNA binding proteins, entry of a virus into a cell, signal induction by neurotransmitters, cellular recognition and so on. The design of artificial, abiotic systems capable of displaying processes of highest efficiency and selectivity requires the correct manipulation of the energetic and stereochemical features of the non-covalent, intermolecular forces (electrostatic interaction, hydrogen bonding, van der Waals forces, etc.) within a defined molecular architecture. In doing so, the chemist finds inspiration in the ingenuity of biological events and encouragement in the demonstration that such high efficiencies, selectivities, and rates can indeed be attained. However, chemistry is not limited to systems similar to those found in biology, but is free to create unknown species and to invent novel processes.

Supramolecular chemistry is a highly interdisciplinary field of science covering the chemical, physical, and biological features of the chemical species of greater complexity than molecules them-

The chemist finds inspiration in the ingenuity of biological events. However, chemistry is not limited to systems similar to those found in biology.

Supramolecular chemistry is a highly interdisciplinary field of science at the intersection of chemistry, biology and physics.

selves, that are held together and organized by means of intermolecular (non-covalent) binding interactions. This relatively young area has been defined, conceptualized, and structured into a coherent system. Its roots extend into organic chemistry and the synthetic procedures for molecular construction, into coordination chemistry and metal ion-ligand complexes, into physical chemistry and the experimental and theoretical studies of interactions, into biochemistry and the biological processes that all start with substrate binding and recognition, into materials science and the mechanical properties of solids. A major feature is the range of perspectives offered by the cross-fertilization of supramolecular chemical research due to its location at the intersection of chemistry, biology, and physics. Drawing on the physics of organized condensed matter and expanding over the biology of large molecular assemblies, supramolecular chemistry expands into a *supramolecular science*. Such wide horizons are a challenge and a stimulus to the creative imagination of the chemist. Thus, supramolecular chemistry has been rapidly expanding at the interfaces of chemical science with physical and biological phenomena.

The emergence of any novel field of science is linked to the past. Where would the roots of supramolecular chemistry reach? It is Paul Ehrlich who recognized that molecules do not act if they do not bind ("*Corpora non agunt nisi fixata*") thus introducing the concept of *receptors*. But binding must be selective, a notion that was enunciated by Emil Fischer in 1894 and very expressively presented in his celebrated "lock and key" image of steric fit, implying geometrical complementarity, that lays the basis of molecular recognition. Finally, selective fixation requires interaction, affinity between the partners, that may be related to the idea of *coordination* introduced by Alfred Werner, supramolecular chemistry being in this respect a generalization of coordination chemistry.

With these three concepts, fixation, recognition and coordination, the foundations of supramolecular chemistry are laid.

> The foundations of supramolecular chemistry are laid with three concepts: fixation, recognition and coordination.

"Just as there is a field of molecular chemistry based on the covalent bond, there is a field of supramolecular chemistry, the chemistry of molecular assemblies and of the intermolecular bond". It is "chemistry beyond the molecule", whose objects are "supramolecular entities, supermolecules possessing features as well defined as those of molecules themselves".

Supramolecular chemistry started with the selective binding of alkali metal cations by natural as well as by synthetic macrocyclic and macropolycyclic ligands, the crown ethers and cryptands. This led to the emergence of *molecular recognition* as a new domain of chemical research that expanded over other areas and became supramolecular chemistry. It underwent explosive growth with the development of synthetic receptor molecules of numerous types for the strong and selective binding of cationic, anionic or neutral complementary substrates of organic, inorganic or biological nature, by means of various interactions (electrostatic, hydrogen binding, van der Waals, donor-acceptor). Molecular recognition implies the (molecular) storage and (supramolecular) retrieval of molecular structural information.

Many types of receptor molecules have already been explored (crown ethers, cryptands, spherands, cavitands, calixarenes, cyclophanes, cryptophanes, etc.). Still many others may be imagined for the binding of complementary substrates of chemical or biological significance, for instance for the development of substrate specific sensors or for the recognition of structural features in biomolecules (nucleic acid probes, affinity cleavage reagents, enzyme inhibitors, etc.).

The combination of recognition features with reactive functions generates *supramolecular reagents and catalysts* that operate in processes involving two main steps: substrate recognition followed by its transformation into products. Because of their relationship with enzymatic catalysis, they present protoenzymatic and biomimetic features. By nature they are abiotic reagents, that may perform the same overall processes as enzymes without following

Supramolecular chemistry started with the selective binding of alkali metal cations by natural as well as by synthetic macrocyclic and macropolycyclic ligands, the crown ethers and cryptands.

the same mechanistic pathways. More importantly, they may also effect highly efficient and selective reactions that enzymes do not perform. This represents a very important area for further development, that may lead to a range of reactive receptor molecules combining substrate specificity with high reactional efficiency and selectivity. Much work remains to be done that should contribute very significantly to the understanding of chemical reactivity and to its application in industrial processes.

Suitably modified receptors act as *carriers* for the selective *transport* of various types of substrates through artificial or biological membranes. Again, many further developments may be envisaged, concerning for instance the construction of selective membrane sensors or the transport of drugs through biological barriers which may include targeting if suitable target-selective recognition groups are introduced.

Figure 2 From molecular to supramolecular chemistry: molecules, supermolecules, molecular and supramolecular devices.

Recognition, reactivity and transport represent the three basic functional features of supramolecular species. (*Figure 2*).
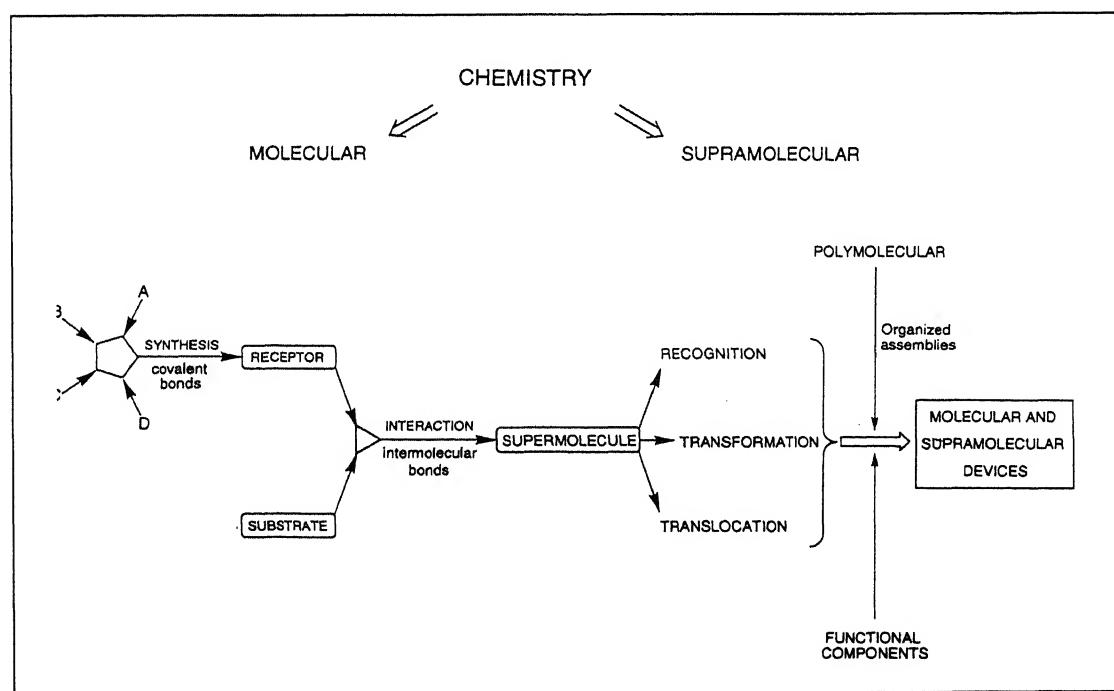
A further important line of development concerns the design of *supramolecular devices* built on photoactive, electroactive or ionoactive components, operating respectively with photons, electrons and ions. Thus, a variety of photonic devices based on photoinduced energy and electron transfer may be imagined. Molecular wires, ion carriers and channels facilitate the flow of electrons and ions through membranes. Such functional entities represent entries into molecular photonics, electronics and ionics, that deal with the storage, the processing and transfer of materials, signals and information at the molecular and supramolecular levels.

A whole field, at the interface with physics, microelectronics and microoptics lies here, which has barely been explored and remains wide open, presenting such intriguing goals as storage (battery), amplification, switching, rectification, etc., devices. The chemistry of molecular signal generation, processing, transfer, conversion and detection, *semiochemistry*, touches upon both physical and biological signalization processes.

The most recent developments concern the implementation of *molecular information* and recognition as a means of controlling the evolution of supramolecular species as they build up from their components. Thus, beyond the preorganization used in the construction of molecular receptors, lies self-organization. It involves the design of systems capable of spontaneously generating well-defined supramolecular entities by self-assembling from their components in a given set of conditions.

The information necessary for the process to take place and the programme that it follows must be stored in the components and they operate via an algorithm based on molecular recognition events. Thus, these systems may be termed *programmed supramolecular systems*.

Self-assembly and *self-organization* have recently been implemented in several types of organic and inorganic systems. By clever use of metal coordination, hydrogen bonding or donor-acceptor in-

> Molecular wires, ion carriers and channels represent entries into molecular photonics, electronics and ionics, that deal with the storage, the processing and transfer of materials, signals and information.

teractions, researchers have achieved the spontaneous formation of a variety of novel and intriguing species such as inorganic double and triple helices, termed helicates, catenanes, threaded entities (rotaxanes), cage compounds, etc. For instance, by the self-assembly of eleven particles, five ligands of two different types and six copper (I) metal ions, a closed, cage-like structure has been obtained spontaneously and selectively in one stroke!

A further major development along these lines, concerns the design of molecular species displaying the ability to form by *self-replication*. This has been realized using components containing suitable recognition groups and reactive functions.

In a study of helicate self-assembly from a mixture of different ligands and different metal ions, it has been found that only the 'correct' helical complexes are formed through self-recognition. In a broader perspective, this points to a change in paradigm from pure compounds to instructed mixtures, that is from seeking chemical purity to designing programmed systems composed of mixtures of instructed components capable of spontaneously forming well-defined superstructures. One may venture to predict that this trend will represent a major line of development of chemistry in the years to come: the spontaneous but controlled build up of structurally organized and functionally integrated supramolecular systems from a preexisting 'soup' of instructed components following well-defined programs and interactional algorithms. Thus, the study of self-processes represents an area of rapidly increasing activity.

In addition to dealing with the oligomolecular supermolecules, well-defined species resulting from the specific intermolecular association of a few components, supramolecular chemistry deals also with polymolecular assemblies formed by the spontaneous association of a large number of components into a specific phase (films, layers, membranes, vesicles, micelles, mesophases, surfaces, solids, etc.). There lies here a vast and fertile domain of research. Molecular recognition between complementary compo-

A space-filling model of this supramolecule is shown on the cover page. The picture presents the axial view of the *self-assembled* cylindrical cage-like structure. The ligands shown in blue, one on the top and one on the bottom, each have three bipyridine-like units for coordination. Three other rod-shaped ligands, shown in yellow, have two bipyridine units each. Each copper (I) ion is coordinated by two bipyridine units, one from the *blue* ligand, and one from the *yellow* ligand. Thus, the five ligands together hold six copper ions to produce this cylinder.

Designing programmed systems composed of mixtures of instructed components capable of spontaneously forming well-defined superstructures represent s a major line of development of chemistry in the years to come.

nents provides means for directing the architecture of polymolecular assemblies and for endowing them with novel properties, such as for instance the selective binding of substrate molecules to layers and surfaces. It allows the design and engineering of *supramolecular materials*, in particular of liquid crystalline and of polymeric nature. For instance, the recognition-induced self-assembly of complementary components generates liquid crystalline 'polymers' of supramolecular nature. A sort of supramolecular polymer chemistry is thus emerging and receiving more and more attention.

Molecular recognition-directed processes also provide a powerful entry into *supramolecular solid state chemistry* and crystal engineering. The ability to control the way in which molecules associate may allow the designed generation of desired architectures in the solid state. Modification of surfaces with recognition units could lead to selective surface binding and to recognition-controlled adhesion.

The design of molecular information controlled 'programmed' systems represents new horizons in materials engineering and processing towards 'smart', functional supramolecular materials, such as self-assembling nanostructures, organized and functional species of nanometric dimensions that define a supramolecular nanochemistry.

It has become clear that the keyword of supramolecular chemistry is not size but information. Supramolecular species spontaneously build up from their components and accomplish complex tasks on the basis of the encoded information and instructions. Thus, if sizewise "there's plenty of room at the bottom" as the celebrated aphorism of Richard Feynman goes, through supramolecular chemistry "there's even more room at the top"!

In chemistry, like in other areas, the language of information is extending that of constitution and structure as the field develops towards more and more complex architectures and behaviour. And supramolecular chemistry is paving the way towards com-

> Supramolecular species spontaneously build up from their components and accomplish complex tasks on the basis of the encoded information and instructions.

prehending chemistry as an *information science*. In the one hundred years since 1894, molecular recognition has evolved from Emil Fischer's "lock and key" image of the age of mechanics towards the *information paradigm* of the age of electronics and communication. This change in paradigm will profoundly influence our perception of chemistry, how we think about it, how we perform it. Instructed chemistry extends from selectivity in the synthesis and reactivity of molecular structures to the organization and function of complex supramolecular entities. The latter rely on sets of instructed components capable of performing on mixtures, specific operations that will lead to the desired substances and properties by the action of built-in self-processes.

Supramolecular chemistry has started and developed as defined by its basic object, the chemistry of the species generated by non-covalent interactions. Through recognition and self-processes it has led to the concepts of (passive and active) information and of programmed systems, becoming progressively the chemistry of molecular information, its storage at the molecular level, its retrieval, transfer and processing at the supramolecular level.

The outlook of supramolecular chemistry is toward a general science of informed matter, bringing forward in chemistry the third component of the basic trilogy matter-energy-information.

> Molecular recognition has evolved from Emil Fischer's "lock and key" image of the age of mechanics towards the *information paradigm* of the age of electronics and communication.

The progression from elementary particles to the nucleus, the atom, the molecule, the supermolecule and the supermolecular assembly represents steps up the ladder of *complexity*. Particles interact to form atoms, atoms to form molecules, molecules to form supermolecules and supramolecular assemblies, etc. At each level novel features appear that did not exist at a lower one. Thus a major line of development of chemistry is towards complex systems and the emergence of complexity.

The highest level of complexity is that expressed in that highest form of matter, living matter, life, which itself culminates in the

The challenge for chemistry lies in the development of non-natural systems displaying desired structural features and carrying out functions *other* than those present in biology with (at least) comparable efficiency and selectivity.

brain, the plasticity of the neural system, epigenesis, consciousness and thought.

Chemistry and notably supramolecular chemistry entertain a double relationship with biology. Numerous studies are concerned with substances and processes of biological or *biomimetic* nature. There has been a profound evolution by which the chemist appropriates and diverts the power of the natural chemical processes of biology to the goals of chemistry, for instance, in the use of enzymes as reagents, the generation of catalytic antibodies, the control of gene expression, the development of molecular diversity techniques, etc. Conversely, the scrutinization of biological processes by chemists has provided understanding on a precise molecular basis and ways for acting on them by means of suitably designed substances. Thus, the cultures of chemistry and biology are intimately linked and coming closer and closer together.

On the other hand, the challenge for chemistry lies in the development of *abiotic*, non-natural systems, figments of the imagination of the chemist, displaying desired structural features and carrying out functions *other* than those present in biology with (at least) comparable efficiency and selectivity. Subject only to the constraints one chooses to impose and not to those of the living organism, abiotic chemistry is free to invent new substances and processes. The field of chemistry is indeed broader than that of the systems actually realized in nature.

Figure 3 A comparison between chemistry and biology with respect to the two parameters : complexity and diversity/breadth.
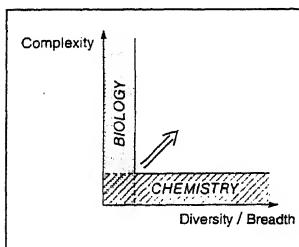
The future path of chemistry will be shaped by both inside and outside forces. Its evolution towards increasing diversity and towards increasing complexity also takes biological phenomena as points of reference. The specificity of chemistry may be stressed by comparing biology and chemistry with respect to these two basic parameters, *complexity* and *diversity*. As presented in *Figure 3*, biology is of extreme complexity, however the substances on which it is based belong to defined classes, and although tremendously rich are nevertheless limited in variety of types.

Chemistry, on the other hand, is still of very low complexity compared to biology but its breadth, the diversity of its substances, is infinite, being limited only by the imagination of the chemist in endlessly combining and recomposing the basic bricks of chemical architectures, thus filling in the unlimited white area in the complexity-diversity diagram.

The chemist finds illustration, inspiration and stimulation in natural processes, as well as confidence and reassurance since they are proof that such highly complex systems can indeed be achieved on the basis of molecular components. One might say that science, notably chemistry, relies on the biological world through an existence axiom; the mere fact that biological systems and, in particular, we human beings exist, demonstrates the fantastic complexity of structure and function that the molecular world can present; it shows that such a complexity can indeed exist despite our *present* inability to understand how it operates and how it has come about. So to say, if we did not exist we would be unable to imagine ourselves! And the molecular world of biology is only one of all possible worlds of the universe of chemistry, that await to be created at the hands of the chemist.

With respect to the *frontiers of life* itself three basic questions may be asked: How? Where? Why?

The first concerns the origin of life on earth as we know it, of our biological world. The second considers the possibility of extraterrestrial life, within or beyond the solar system. The third question wonders why life has taken the form we know; it has as corollary the question whether other forms of life can (and do) exist: is there 'artificial life'?; it also implies that one might try to set the stage and implement the steps that would allow, in a distant future, the creation of artificial forms of life.

Such an enterprise, which one cannot (and should not) at the present stage outline in detail except for initial steps, rests on the presupposition that there may be more than one, several expres-

Chemistry is also an art in its very essence, by its ability to invent the future and to endlessly recreate itself.

sions of the processes characterizing life. It thus invites one to the exploration of the *frontiers of other lifes* and of the *chemical evolution* of living worlds.

Questions have been addressed about which one may speculate, let one's imagination wander, perhaps even set paths for future investigations. However, where the answers lie is not clear at present and future chemical research towards ever more complex systems will uncover new modes of thinking and new ways of acting that we at present do not know about and may even be unable to imagine.

The perspectives are definitely very (too?) wide and it will be necessary to distinguish the daring and visionary from the utopic and illusory! On the other hand, we may feel like progressing in a countryside of high mountains: the peaks, the goals are visible and identifiable or may become so as progress is made, but we do not yet know how to reach them. We may find landslides, rockfalls, deep crevices, tumultuous streams along the way, we may have to turn around and try again, but we must be confident that we will eventually get there. We will need the courage to match the risks, the persistence to fill in the abyss of our ignorance and the ambition to meet the challenges, remembering that *"who sits at the bottom of a well to contemplate the sky, will find it small"* (*Han Yu*, 768-824).

But to chemistry, the skies are wide open, for if it is a science, it is also an art. By the beauty of its objects, of course, but also in its very essence, by its ability to invent the future and to endlessly recreate itself.

The chemist creates original molecules, new materials and novel properties from the elements provided by nature — indeed entire new worlds.

Like the artist, the chemist engraves into matter the products of creative imagination. The stone, the sounds, the words do not contain the works that the sculptor, the composer, the writer express from them. Similarly, the chemist creates original molecules, new materials and novel properties from the elements provided by nature, indeed entire new worlds, that did not exist

before they were shaped at the hands of the chemist, like matter is shaped by the hand of the artist.

Indeed chemistry possesses this creative power as stated by *Marcelin Berthelot "La chimie crée son objet". ("Chemistry creates its object")*. It does not merely fabricate objects, but creates its own object. It does not preexist, but is invented as progress is made. It is not just waiting to be discovered, but it is to be created.

The essence of chemical science finds its full expression in the words of that epitome of the artist-scientist *Leonardo da Vinci: "... dove la natura finisce di produrre le sue spezie, l'uomo quivi comincia con le cose naturali, con l'aiutorio di essa natura, a creare infinite spezie...".* (*"Where nature finishes producing its own species, man begins, using natural things and with the help of this nature, to create an infinity of species..."*).

The essence of chemistry is not only to discover but to invent and, above all, *to create*. The book of chemistry is not only to be read but to be written! The score of chemistry is not only to be played but to be composed!

> The essence of chemistry is not only to discover but to invent and to create.

> Transcript of the Third Rajiv Gandhi Science and Technology Lecture delivered at Hyderabad on 22 December 1995.

Address for correspondence
Jean-Marie Lehn
Collège de France, Paris and
Université Louis Pasteur,
Strasbourg, France.

"To start the Universe, light the fuse and turn the clock"

AYAN GUHA
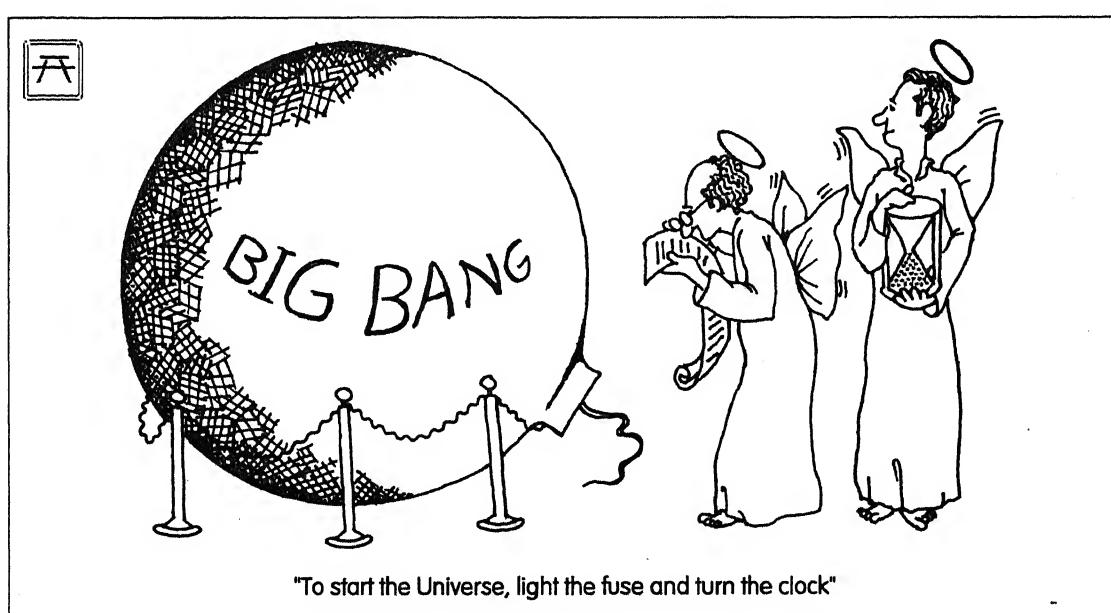
The Earth's Changing Climate

Man-Made Changes and Their Consequences

P K Das

Various human activities have contributed to an increase in the levels of greenhouse gases. The contribution that these make to global warming requires further investigation, because there are several negative feedback mechanisms which inhibit the warming process.

P K Das
is a former Director
General of the Meteoro-
logical Department of
India. After retiring in
1983, he taught Meteorolo-
gy at the University of
Nairobi in Kenya (1983-85)
and later at the Indian
Institute of Technology
(IIT) in New Delhi (1985-
89). Subsequently, he was
Scientist Emeritus of
CSIR from 1989 to 1992.
He is the author of two
books on the monsoons
and a pioneer in numerical
weather prediction
in India.

Introduction

Many generations of scientists have been intrigued by the forces that change our climate. This is more so today because of the need for sustainable development. Sustainable development implies a growth strategy that will not harm the environment.

In this article we will be mainly concerned with anthropogenic, or man-made causes of climate change, because a review of the entire climate system is not possible within the ambit of a short article. We will focus on the atmosphere, although a brief reference will be made to the role of oceans. Global warming by anthropogenic emissions is referred to as the greenhouse effect because the atmosphere acts as a greenhouse. It allows solar radiation to pass through but traps the outgoing infrared radiation from the earth. Realizing the importance of global warming, the World Meteorological Organization (WMO) and the United Nations Environmental Programme (UNEP) have set up an Intergovernmental Panel on Climate Change (IPCC). We will refer to their assessment reports.

In 1861, Dr John Tyndall, the English physicist, was the first to draw our attention to the rapid absorption of heat by humid air in the atmosphere. A little later, in 1896, the Swedish chemist Arrhenius presented a paper on the sensitivity of the atmosphere

to small changes in its composition. These perceptive papers were the earliest to stress the greenhouse effect.

The Earth's Radiation Budget

The average solar energy intercepted by a unit area of the earth's surface is S/4, where S is the solar constant. The factor of 1/4 allows for the radiation falling at an angle on large parts of the earth's surface. As S is around 1376 watts per square metre (W/m^2), S/4 is 344 W/m^2. About one-third of this is either back-scattered by air molecules or reflected back to space and this fraction is often called the planetary albedo. So the amount entering the top of the atmosphere is about 240 W/m^2. Solar radiation is in the short wavelength range between 0.2 and 4.0 microns (μm).

The earth, in turn, radiates as a black body. It obeys the Stefan-Boltzmann law and its emission is proportional to the fourth power of the earth's surface temperature. This infrared radiation from the earth is in the long wavelength range from 4.0 to 80 μm. Many atmospheric gases have a rotation as well as a vibration-rotation spectrum, so they are able to absorb and also emit radiation. This has two interesting consequences. First, the net upward emission by atmospheric constituents, clouds and a small part of the earth's emission adds up to 240 W/m^2. This represents a balance between incoming and outgoing radiation. Secondly, there is a downward counter radiation from the atmosphere towards the earth. This traps the earth's infrared radiation like a greenhouse. The radiation budget is shown in *Figure 1*.

It is interesting to note that in the absence of a greenhouse effect, the earth's mean temperature would have been 255°K, instead of the observed value of 288°K. Greenhouse warming is thus about 33°K. This is much more pronounced around the planet Venus, whose atmosphere contains over 90% carbon dioxide (CO_2), a prominent greenhouse gas. Without a greenhouse effect the mean surface temperature of Venus would have been 227°K, but greenhouse warming raises it to 750°K. This represents warming by 523°K!

The earth radiates as a black body. It obeys the Stefan-Boltzmann law and its emission is proportional to the fourth power of the earth's surface temperature.

The net upward emission by atmospheric constituents, clouds and a small part of the earth's emission adds up to 240 W/m^2.

TOTAL SOLAR RADIATION 344 W/m^2

SUN

NET INCOMING SOLAR RADIATION 240 W/m2
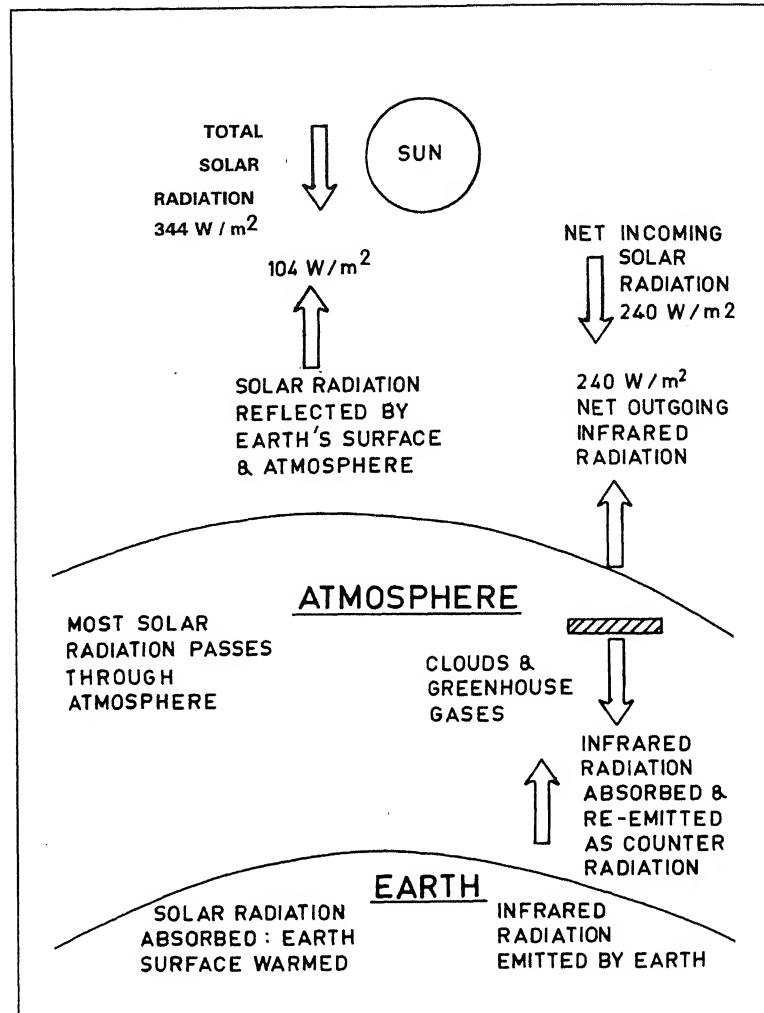
104 W/m^2

SOLAR RADIATION REFLECTED BY EARTH'S SURFACE & ATMOSPHERE

240 W/m^2 NET OUTGOING INFRARED RADIATION

ATMOSPHERE

MOST SOLAR RADIATION PASSES THROUGH ATMOSPHERE

CLOUDS & GREENHOUSE GASES

INFRARED RADIATION ABSORBED & RE-EMITTED AS COUNTER RADIATION

EARTH

SOLAR RADIATION ABSORBED : EARTH SURFACE WARMED

INFRARED RADIATION EMITTED BY EARTH

It is interesting to note that in the absence of a greenhouse effect, the earth's mean temperature would have been 255°K, instead of the observed value of 288°K.

The principal greenhouse gases in our atmosphere are: (i) Carbon dioxide (CO_2), (ii) Methane (CH_4), (iii) Nitrous oxide (N_2O), (iv) Ozone (O_3) and (v) halocarbons.

Halocarbons are compounds of carbon and halogens (chlorine, bromine and fluorine). Some refer to them as the chlorofluorocarbons, or the CFC's. Of late, another class of carbon compounds, the hydrogenated CFC's (HCFS) are also in the category of greenhouse gases. We will refer to them collectively as the halocarbons. Halocarbons are inert gases with a long life of 50 to 100

years. They are injected in the lower atmosphere, the troposphere, through refrigerants and aerosol sprays. The halocarbons destroy ozone in the lower stratosphere through chemical action, and as ozone protects us by absorbing harmful ultraviolet radiation from the sun, serious efforts are now in progress to replace halocarbons by other eco-friendly chemicals.

Carbon dioxide, the principal greenhouse gas, is released into the atmosphere by burning fossil fuels, and by the cement manufacturing industry. But, it is removed from the atmosphere by the photosynthesis of plants. The largest reservoirs of carbon are in the deep oceans. Some of this reaches the atmosphere when waters from the deep ocean are brought up to the surface.

The main sources of methane (CH_4) are paddy fields and natural wetlands. Intestinal fermentation in animals also generates methane. Of the other greenhouse gases, nitrous oxide (N_2O) is obtained partly from the oceans and partly from fertilizers that are used in agriculture. The burning of biomass also generates N_2O.

Ozone (O_3) is a natural constituent of the atmosphere. Depending on the latitude, the ozone content in a vertical column of the atmosphere shows a peak concentration in the middle and lower stratosphere (15 to 35 km). This is known as the ozone layer. There is a gradual increase in ozone concentration from the equator and the tropics to higher latitudes. A matter of great concern is the recent discovery of a sharp decrease in ozone over Antarctica, especially during spring. The observations suggest a 50% decrease in 1987 when compared to values prevailing ten years ago.

A disturbing feature is a gradually increasing trend for CO_2, CH_4, N_2O and halocarbons. The rates of increase range from 0.25 to 0.9% per year for CO_2, CH_4 and N_2O, but it is as high as 4% for the halocarbons. There is some uncertainty about methane because it has shown a decreasing trend very recently. But, the overall need today is to decrease the emission rates of greenhouse gases. This

is because of their adverse impact on climate. In this context, it is relevant to ascertain their contributions to global warming.

Radiative Forcing and Global Warming Potential (GWP)

The change in radiation flux at the tropopause (*Figure 2b*) caused by the absorption of either solar or infrared radiation is defined as the 'radiative forcing'. Positive radiative forcing leads to warming, while a negative value cools the earth's surface. Radiative forcing could be either direct or indirect. Direct forcing occurs when the concerned gas is involved, but when the gas interacts with another gas by chemical interaction it leads to indirect forcing.

The hydroxyl (OH) radical plays an important role in indirect radiative forcing because it is a strong oxidizing agent. There are several reactions which lead to the formation of the OH radical. The OH radical is formed, for example, when solar ultraviolet radiation of wavelength less than 3 μm reacts with ozone. This gives rise to electronically excited oxygen atoms which react with water vapour to form hydroxyl radicals.

The IPCC assessments of radiative forcing are shown in *Figure 2*. The confidence for each assessment is shown in the bottom row

---
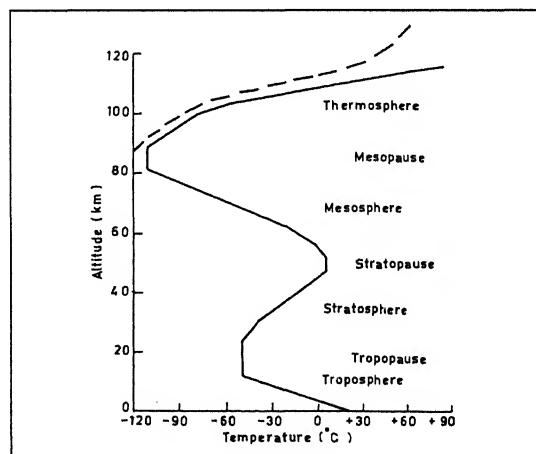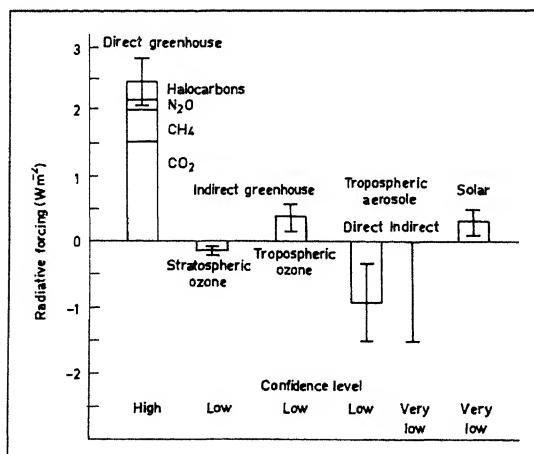
ɔle 1 Global Warming Potential (GWP) for 100 years

Life time (yr.)	GWP
-	1
14.5 ± 2.5	24.5
120	320
50 ± 5	4000
102	8500

:ertainty is around ±35% for each gas *[Source: Radiative forcing of change, report of the Scientific Assessment Group of IPCC, 1994].*

gure. Of some interest are the low values given for solar i ty. This suggests that variations in the solar constant, for ≥ₒ are relatively unimportant.

ϽC report also suggests Global Warming Potential (GWP) ɟex for the greenhouse effect. It represents the cumulative ˉradiative forcing between now and a later time, say, a 100 om now. For the purpose of calculating the GWP the caused by a unit mass of gas is considered, and the index ≥ssed in terms of a reference gas, namely, CO_2. *Table 1* s the GWP index for five principal greenhouse gases.

ɦe uncertainty factor is certainly large, the table indicates ⌐ large warming potential of the halocarbons compared to ≥rs.

ive Feedback Mechanisms

ϽC estimates that the earth's mean temperature has risen Ͻ 0.6°K in the last 100 years. If no action is taken to restrict ʹns, the warming rate is estimated to be 0.3°K for every ten the future, with an uncertainty range between 0.2° and 0.5°K.

ge uncertainty range is due to negative feedbacks which we ʟnderstand entirely. The four principal cooling mecha-

nisms are: (i) ozone depletion in the lower stratosphere, (ii) back-scattering and reflection of short wave solar radiation by dust particles and aerosols, especially in the vicinity of sulphur emissions, (iii) clouds and (iv) CO_2 at high altitudes of the atmosphere. These feedbacks are briefly described in turn.

● *Stratospheric ozone depletion*

Ozone has a dual impact in the lower stratosphere (25 km). Firstly, the depletion of ozone by halocarbons implies less absorption of solar radiation by ozone, due to which more of it reaches the earth. This has a warming effect. But, less ozone also implies less absorption of outgoing infrared radiation. This cools the stratosphere. And, a cooler stratosphere emits less infrared radiation to the troposphere below. The net effect is one of cooling. Around 20-25 km, the cooling process usually dominates. We have already mentioned the ozone hole over Antarctica. The reasons for this are still unclear. Polar stratospheric clouds, which hold a wide variety of particulates and sulphate aerosols also contribute to ozone decline. In addition, chemical interactions release large amounts of chlorine that destroy ozone.

● *Aerosols*

Aerosols also have a dual feedback. They scatter and reflect solar radiation, especially if they are covered with sulphates. This has a cooling effect. But, they also absorb both solar and infrared radiation. On many occasions the scattering features dominate, and this leads to cooling. Volcanic eruptions increase the dust load of the atmosphere and this increases negative feedback. The recent eruption of Mount Pinatubo was estimated to have decreased the temperature by a few tenths of a degree over many parts of the earth.

● *Clouds*

V Ramanathan, was the first to point out that clouds have a negative feedback. Clouds reflect solar radiation upwards to

The recent eruption of Mount Pinatubo was estimated to have decreased the temperature by a few tenths of a degree over many parts of the earth.

space. This cools the earth's surface, but they also absorb terrestrial infrared radiation which warms the earth. Cooling predominates in the thicker clouds, but the high altitude thin clouds generate more warming. Ramanathan's suggestion is that if global warming leads to warmer sea surface temperatures (SST) in the tropics, the ensuing convection could cause thick clouds. This will cool the atmosphere by a negative feedback of about 16 W/m^2. The clouds will thus act as a thermostat and set an upper limit to the sea surface temperature. This view has not found general agreement, because observations indicate that warm SSTs do not always generate thick clouds. The issue is still open.

● *High altitude* CO_2

In a recent communication to *Nature* (Vol.304, 8 March, 1990) Cicerone from UCLA in the USA suggests that increasing concentrations of CO_2 in the mesosphere (80 to 100 km above sea level) will cause stratospheric cooling because CO_2 molecules are efficient radiators of energy to space through infrared emissions. This negative feedback has not yet received much attention.

The negative feedbacks that we have mentioned above need further study and research. Unless this is done, future forecasts of global warming will remain uncertain.

Natural Variability and Forced Oscillations

One of the difficult tasks that confront atmospheric scientists today is to ascertain if forcing mechanisms, such as global warming, generate responses that are larger than the natural variability of climate.

It is now recognized that on a regional scale global warming can alter the statistical mean or the variance of the distribution of a meteorological variable (See Rajeeva L Karandikar's article 'On Randomness and Probability' in *Resonance* Vol. 1 No. 2). This is shown in *Figures 3(a)* and *3(b)* and, as we can see, this will lead to
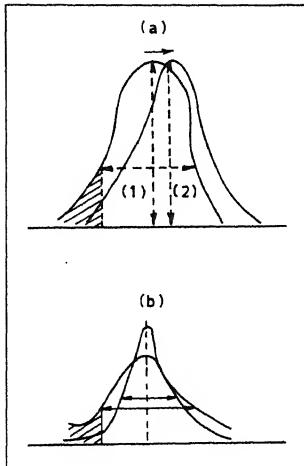
Figure 3 Impact of a change in the mean and variance of a meteorological element due to global warming. Note the enlarged frequency of extreme events.

a change in the frequency of extreme events. Let Figure 3(a) represent the rainfall distribution over a small region, such as a state in our country. The abscissa represents rainfall while the ordinate is the frequency. For simplicity, we will assume a symmetric bell-shaped (normal) distribution. Suppose now that as a result of global warming the median value is changed from (1) to (2), without any change in the shape of the distribution. As we can see from Figure 3(a) the frequency of low rainfall events (droughts) will decrease, while the frequency of high rainfall events will increase leading to more floods. The change in frequency of low rainfall is shown by the hatched area in Figure 3(a). Figure 3(b) shows the outcome of a decrease in the mean rainfall but an increase in the lateral spread or variance of the distribution. In this case there will be an increase in the frequencies of both low rainfall events (droughts) as well as high rainfall events or floods.

But, if we examine the time series of monsoon rainfall, for example, we will observe variations from year to year and decade to decade (*Figure 4*). Can we be sure that these variations are not due to natural causes?

This interesting question was first raised by the late Professor Jule Charney of the Massachusetts Institute of Technology, USA and J Shukla, in the United States. Unfortunately, no clear answers have been found yet. One useful result that has emerged is an inverse relationship between equatorial Pacific SST's and monsoon rainfall. This result has been verified with the help of statistical techniques and mathematical models. The latter now simulate coupled ocean-atmosphere circulations.

If we examine the time series of monsoon rainfall, for example, we will observe variations from year to year and decade to decade. Can we be sure that these variations are not due to natural causes?

In this context it is worth mentioning that a few promising signals of climatic change have been detected in the last decade. They are now attracting a good deal of research, especially for short term climate prediction. These signals are generated by the slow evolution of meteorological variables at the earth's surface. Examples of these variables are sea surface temperature (SST), snow cover and soil moisture.

Annual rainfall (mm)   30 Year Mean (mm)
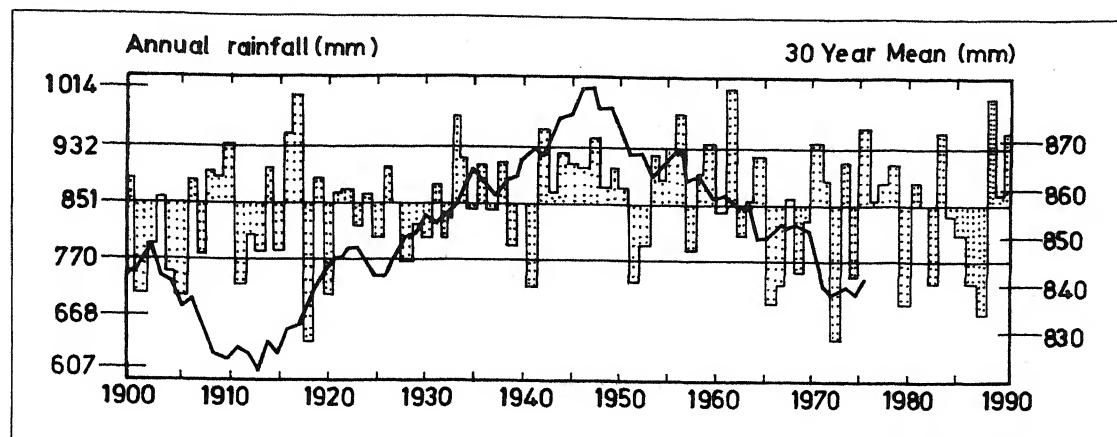
Perhaps the most interesting signal is an ENSO event in the equatorial Pacific. ENSO stands for an El Nino (EN) and the Southern Oscillation (SO). The former, which means 'a child' in Spanish, represents the sudden appearance of warm waters off the coast of Peru. It occurs once every two to seven years. As it occurs around Christmas many refer to it as the 'child Jesus'. It is accompanied by abnormally heavy rain in many parts of the tropics. The fishing industry of Peru suffers heavy damage due to the warm waters of an El Nino. The Southern Oscillation (SO) is the term used to define a see-saw pattern of pressure variations between the Pacific and the Indian Oceans. When pressures are low over the Pacific they tend to be high over the Indian Ocean and vice versa. A Southern Oscillation Index (SOI) measures the intensity and phase of this oscillation. It is the pressure difference between the island of Tahiti in the Pacific and Port Darwin (to represent the Indian Ocean). A negative value of SOI occurs concurrently with an El Nino which is why we refer to them, jointly, as an ENSO event. As an ENSO is associated with weak monsoon rains, its prediction is important for Indian agriculture.

Coupled air-sea models are now used to anticipate SST changes over the equatorial Pacific. Forecasts are being made with lead times ranging from a month to a year. The design of coupled models was pioneered by Zebiak and Cane of the United States in

El Nino represents the sudden appearance of warm waters off the coast of Peru. It occurs once every two to seven years.

A negative value of the Southern Oscillation Index (SOI) occurs concurrently with an El Nino which is why we refer to them, jointly, as an ENSO event. As an ENSO is associated with weak monsoon rains, its prediction is important for Indian agriculture.

1987. To see how the model works, we consider an air-water interface between the atmosphere and the underlying ocean. The governing equations use the observed atmospheric winds to force ocean currents. This leads to variations in SST and they, in turn, determine the pattern of winds over the ocean. As the heat capacity of the ocean is much larger than that of the atmosphere the evolution of the SST variations is a very slow process. The oceans act as the memory of the atmosphere.

While encouraging results have been achieved, much more remains to be done. The performance of the model is very sensitive to the accuracy of the initial conditions. These conditions have to be specified before starting model integration. The Climate Research Centre (CRC) of the National Oceanic and Atmospheric Administration (NOAA) in the USA feels that the skill achieved so far is modest, but further research in this important area is being vigorously pursued.

Is the overall climate predictable? An answer to this intriguing question was the subject of an interesting paper by Professor E N Lorenz of MIT in USA (Tellus, 424, 378-389, 1990). Professor Lorenz has pioneered the modern theory of chaos. He considered a simple model climate consisting of an eastward moving wind current, which is disturbed by waves. Professor Lorenz suggests that short term climatic changes are essentially chaotic. He finds that if the external forcing during winter, for example, is very large then the circulation will become chaotic. And, it could remain chaotic until next spring, even though there might be a decrease in external forcing. This suggests that long range prediction of climate might not be possible. Professor Lorenz's model was a very simple one. It is possible that future models will contain stabilising factors that will improve the limit of predictability. An exciting field of research lies ahead.

Summary and Conclusions

● Anthropogenic emissions of greenhouse gases indicate the possibility of an increase in global warming.

- There are several negative feedbacks which make future projections uncertain at present.
- Clouds often provide a large negative feedback, but they need not make the atmosphere a thermostat.
- It is difficult to separate forced climatic changes from the natural variability of climate.
- Short term climatic changes are mainly chaotic. This could make it difficult to make long range predictions. But, some success can be achieved in predicting variables which evolve slowly, such as, sea surface temperature (SST).

Is the overall climate predictable? Lorenz's experiment suggests that long range prediction of climate might not be possible.

Suggested Reading

J T Houghton, G J Jenkins, J J Ephraums (Ed). Climate change. The IPCC Seventh Assessment, WMO-UNEP. Intergovernmental Panel on Climate Change. Camb. Univ. Press. 1990.

1922 IPCC Supplement, Sc. Assessment of Climate Change, WMO - UNEP. IPCC Report. Geneva.

Radiative Forcing of Climate Change. WMO-UNEP, IPCC Report, Geneva. 1994.

B Bolin, B R Doos, J Jager, R A Warrick. (Ed). The Greenhouse Effect, Climate Change and Ecosystems, SCOPE 29, John Wiley & Sons, Chichester, United Kingdom. 1986.

A Wiin-Nielsen. The Increased Greenhouse Effect: A General Review. *Proc. Indian Acad. Sci.* (Earth Planet. Sci.). 102(1):3-25, 1993.

R E Newell, Zhongxiang Wu. Temperature Limits in the Climate System. *Recent Adv. in Atmos. Sci.*, Indian Nat. Sci. Acad. New Delhi. 1994.

Stephen E Zebiak, Mark A Cane. A Model El-Nino- Southern Oscillation. *Monthly Weather Review.* 115:2262-2278. 1987.

Address for correspondence
P K Das
A-59, Kailash Colony,
New Delhi 110 048. India.

Hermann Hankel said... "In most sciences one generation tears down what another has built, and what one has established another undoes. In mathematics alone each generation adds a new storey to the old structure".

Bertrand Russell wrote... "There was a footpath leading across fields to New Southgate, and I used to go there alone to watch the sunset and contemplate suicide. I did not, however, commit suicide, because I wished to know more of mathematics".

Time in a Timeless Environment

My Life in a Bunker

L Geetha

Geetha L
studies circadian rhythms
in honeybees, mice and
humans. She has had the
rare privilege of being both
an experimental subject as
well as an author of a
scientific study as you will
read in this article.

The fact that every organism possesses an endogenous biological clock can be established when they are made to live in conditions of timelessness. How do human beings perform under such conditions? Human isolation facilities provide such timeless environments where human beings can live comfortably while performing various bodily activities as dictated by their endogenous clock. In this article, I narrate my personal experience in the isolation facility at Madurai Kamaraj University. My stay (on three occasions) has led to the important finding that the menstrual cycle in a human female is not coupled to the sleep-wake cycle. I also describe how such experiments can be useful in the context of shift-working, jet-lag and space studies.

A Bunker?

When you had to prepare for an exam or catch an early flight or train, you probably had the experience of getting up just before the alarm went off. This is possible because of the biological clock that we all possess. Just as we set our watches with standard times from the radio or TV, our body also sets its clock to a 24 hour schedule with the help of cues that are provided by the environment. For instance, the cycle bell of the milkman is sufficient for us to realise that the day has dawned, without even opening our eyes. Other factors such as light, temperature and noise can also provide us with information about time. Almost all physiological parameters in human beings are rhythmic, i.e. they repeat themselves at definite time intervals. For example, the time we go to sleep, the time we wake up, our body temperature, levels of sodium and potassium excretion, water excretion — virtually anything you can think of is rhythmic. All these rhythmic functions take place

Almost all physiological parameters that you think of in human beings are rhythmic i.e they repeat themselves at definite time intervals.

because of the information of time received from the environment we live in.

What would happen if we are made to live in an environment which does not have time cues of any sort? What would happen if we lost track of time? How important is it to have the knowledge of time? All these interesting questions can be answered if we indeed have an environment that completely lacks time cues. In fact such environments have been created in five places in the world! Since these 'time less environments' have been created for conducting experiments on human beings they are called 'facilities'. All five such facilities in the world are rather similar, so let me describe the one in India. Believe me, I can describe it well — I have lived in this timeless facility for a total period of about 100 days (in three bouts)!

Timelessness in the Bunker!

The only isolation facility that we have in India is located in the Department of Animal Behaviour and Physiology, Madurai Kamaraj University (*Figure 1*). The others are located in USA, UK, Switzerland and Japan. The living quarter in the facility is a

Figure 1 *Photograph exhibiting a view of the human isolation chamber.*

25' x 25' room. It is window-less to ensure that daylight does not enter. Artificial lights (fluorescent tubelights with intensity of about 1300 to 1800 lx) are provided for the use of the subject who stays inside. Switching 'on' and 'off' of these lights is left to the discretion of the subjects. The walls of the room are double-layered with sand in between so that no outside noise can be heard. An ambient temperature of about 25° C is maintained throughout the experimental period. For ventilation purposes cool air is passed into the room through a duct with sound muffles. A kitchenette and a bathroom with toilet are attached to the living room. All facilities such as refrigerator, video cassette player, tape recorder, an ergometer for exercising purposes, tables, chairs and materials for cooking are provided in the chamber. The isolation facility is devoid of potential *zeitgebers*[1] (time-givers), viz. clocks, radios, TV, current periodicals etc. The requirements of the subject who stays inside are placed in the antechamber adjoining the isolation facility. Communication with the outside world is in the form of written notes. The facility is manned round the clock from outside. In case of power failures, a generator is switched on within minutes. During the period of stay, all the requirements of the subject are taken care of almost instantaneously! In short, the subject receives 'royal treatment' throughout the duration of the stay!

[1] The cues that give the information of time to the organisms so that they are synchronised with the oscillator.

What Happens in the Bunker?

A number of interesting experiments have been performed in the isolation facilities. Juergen Aschoff and R Wever from Germany performed the pioneering experiments on human circadian rhythms[2]. The experiments conducted in the isolation facility (which is now defunct) in Germany demonstrated for the first time that human beings continue to show rhythms even in the absence of time cues and this rhythmicity is endogenous and dependent on internal clocks. This was done by measuring the sleep-wake cycle (the times of going to sleep and waking up every day) under periods of isolation. Since, we wake up almost at the same time each day under normal conditions, the time difference between the wake-up time of two successive days would be almost

[2] Self-sustained biological rhythms which repeat once in about 24 hours.

24 hours i.e. it is entrained (synchronised) to the 24 hour light-dark cycle of nature. Under periods of isolation, this synchronisation no longer prevails and the subjects are found to drift from the normal 24h cycle, and start free-running[3]. So their sleep-wake cycle would have a period longer than 24 hours under isolation. Aschoff and Wever also proved that for humans, social cues can bring about entrainment [4] to 24 hours thus proving that social cues are more important than light-dark (LD) cycles.

Following these findings, a number of similar experiments have been performed all over the world. Our isolation chamber in India also boasts of a number of interesting findings. Experiments have been carried out here since 1987 when the first subject G Marimuthu, a rhythm researcher himself, entered the isolation chamber in Madurai. During my first visit to the Department of Animal Behaviour, where they specialise in chronobiology (the study of biological rhythms), I happened to have a glimpse of it and instantly, wished to be the next subject. Fortunately for me, the head of the Department, M K Chandrashekaran, who has built this 'one-of-a-kind-chamber', almost immediately agreed to make me the next experimental subject. Since I was involved in these experiments, let me give you a detailed account of how these experiments are generally performed, along with my personal experiences. There was a short spell of pre-isolation period, when I was asked to note down the times of going to sleep and waking up. I also had my core body temperature measured every 6 minutes by using a device called solicorder which records the temperature in a computer chip.

There is a reason behind measuring the body temperature. Though human beings are supposed to be homeotherms, there is a 2° C variation in body temperature every day. Temperature rises to a maximum around midday and dips to a minimum when we are in deep sleep (*Figure 2*). Since the time at which we go to sleep is almost the same every day, the time at which this temperature minimum occurs would also be at approximately 24-hour intervals. Therefore, the temperature cycle (the time interval between

[3] The state of the rhythm in constant conditions in the absence of any time cues where the endogenous period is exhibited.

[4] Synchronization of a self-sustaining rhythm with the period of the imposed *zeitgeber*.

Under periods of isolation, subjects are found to drift from the normal 24h cycle, and start free-running. So their sleep-wake cycle would have a period longer than 24 hours under isolation.
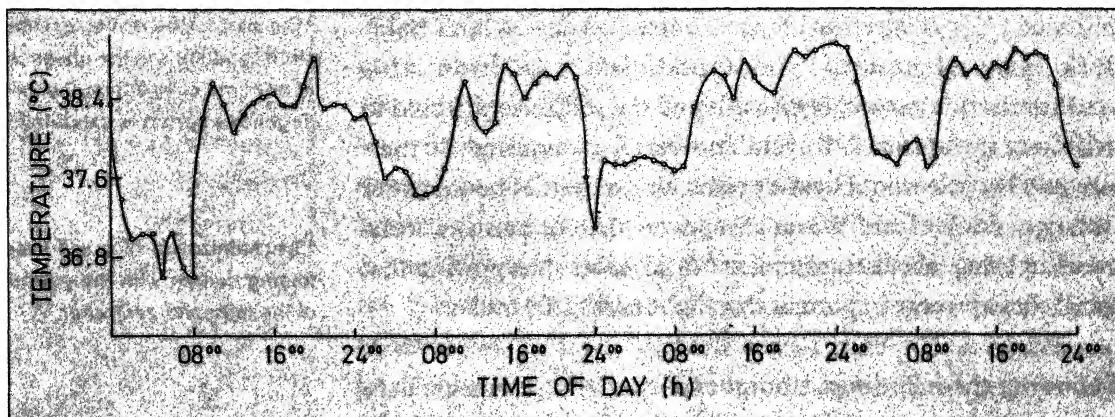
Figure 2 Graph showing the variations in temperature during the course of the day. Each 24 hours in the graph represents one day. Data collected during pre-isolation period for four days has been plotted here. Minimum temperature occurred when I was in deep sleep and maximum temperature almost during midday.

the occurrence of temperature minima of two successive days) would also be about 24 hours. Since we wanted to know whether this temperature rhythm would exhibit any change under isolation, my core body temperature was also measured during pre-isolation, isolation and post-isolation periods. One day prior to my isolation, I was asked to spend the night in the chamber just to get acquainted with the environment and to see if there were any problems which had to be rectified. I entered the isolation chamber on 4th May, 1989 for the first time. I happened to be the first female subject to go in and you can imagine the publicity that I received! All my friends jestingly remarked that I chose the month of May to stay in the chamber to evade the heat of the Madurai sun!

My Life in the Bunker

Thus, my stay in the isolation chamber commenced around 5 p.m. with an august gathering seeing me off! Personally, for me it was a dream coming true! All along I had wanted to be alone somewhere! Never did I dream that, that 'somewhere' was going to be a gorgeous place like the isolation chamber and that I would live like a princess there. We had done an enormous amount of 'shopping' for my stay the previous evening and I essentially spent the first few hours unpacking and organising. A kind of peace that I had never experienced before in my life enveloped me when I went to bed on the first day of my isolation. The next morning I opened my eyes to look for the time in the clock and with a jolt

remembered that this was the isolation chamber and that I would not know the time for quite some time! Gradually, I got used to my life of loneliness and timelessness in the chamber. It was indeed a unique experience in which there was no time restriction to do anything. There was no need to wake up, go to sleep, eat or do anything at a particular time and there was nobody to give me exhortations of any sort. It was a life that I lived for myself and I felt great about it. Incidentally, I feel this was one situation where we can realize how much we comply with the needs of society and make compromises!

In spite of the lack of restrictions, I was somewhat organised although I performed all the activities, whenever they suited me! There is a panel of 20 buttons on one of the walls, each one corresponding to a particular function. For example, No. 1 corresponds to 'wake up', No. 2 for 'out of bed' and so on. I was supposed to activate these buttons as and when I performed these functions which got recorded outside in a device called an 'event recorder'. Thus, at a given time, the people outside knew what I was doing. For example, my sleep-wake cycle was monitored from outside by finding out when I was going to sleep and when I was waking up. I also pressed the button on channel No. 8, whenever I thought 2 hours had elapsed. This is called 'time estimation' to see how accurately a subject can measure the passage of time in a time-less environment. My ambulatory movements were measured by an activity-monitor which I wore on my left hand. With all these paraphernalia, I was quite a sight, but it did feel great to look different. Only, there was no one to look at me!

Apart from the 'pressing of buttons' I did not have anything much to do for the sake of the experiment. The whole day was mine to spend as I wished. I spent my time reading, watching movies, listening to music and generally relaxing and enjoying myself. Of course, at times, I also read some chronobiology. The rest of the time, I spent in trying out new recipes. Generally, the subjects are asked to cook their own food because of the problems encountered in providing food from outside. For example, the first subject who

It was indeed a unique experience in which there was no time restriction to do anything.

My sleep-wake cycle was monitored from outside by finding out when I was going to sleep and when I was waking up.

A few minutes after I came out of the chamber, I was asked to guess the date. I said, much to the amusement of people around, that it was 26th May when in reality, it was 8th of June!

entered isolation demanded *idlis* when it was 3 AM in the outside world and it took enormous efforts for the outsiders to arrange for *idlis* at that time. As you realise, they could not even tell him that *idlis* were not easily available at that time of the day. Dealing with the psyche of the subject inside is a tricky affair, because even small failures might lead to giving some inadvertent time cues. Hence, from then on, the subjects were asked to prepare their own food. Usually all the subjects who entered (7 people had lived in it before me) lost weight. I had visions of a 'very slim self' coming out of the chamber after the experiment, and in fact it did happen! I had lost 5 kgs on the whole! The reason for this was known to me only after the experiment was over and I came out of the chamber. In fact, it was quite a surprise (and a shock!) for me when I was told that the experiment was over and that I had to come out. This is because, originally, it was agreed that I would live in the chamber for at least a month. But, at the time they called me out, I had counted only 22 days! I was quite cross with the people outside for abruptly stopping the experiment after 22 days. A few minutes after I came out of the chamber, I was asked to guess the date. I said, much to the amusement of people around, that it was 26th May when in reality, it was 8th of June! I had lost 13 precious days of my life. But I feel, it was all worth it! In short, I had spent 35 calendar days as 22 subjective days (subjective day is the time that I presumed to be one day as opposed to the calendar day). My rhythm started drifting every day after my entry and my period of wakefulness kept increasing day by day and so did my duration of sleep. Thus, I had been awake for a maximum of 34 hours and had slept for a maximum of 19 hours! As a consequence, the duration of my one day inside was on an average 45.9 hours. This is called 'circa-bidian' i.e. about 48 hours per day. So my sleep-wake cycle had a rhythm of about 45.9 hours (*Figures 3, 4*). Interestingly, my temperature rhythm still maintained about 24-hour rhythmicity, i.e. temperature minima continued to occur once in about 24 hours despite the variation in the sleep-wake cycle. My body temperature started coming to a minimum twice a day (once when I was awake and should have gone to sleep and once more when I was actually asleep). As a result, I had two temperature minima

My rhythm started drifting every day after my entry and my period of wakefulness kept increasing day by day and so did my duration of sleep.
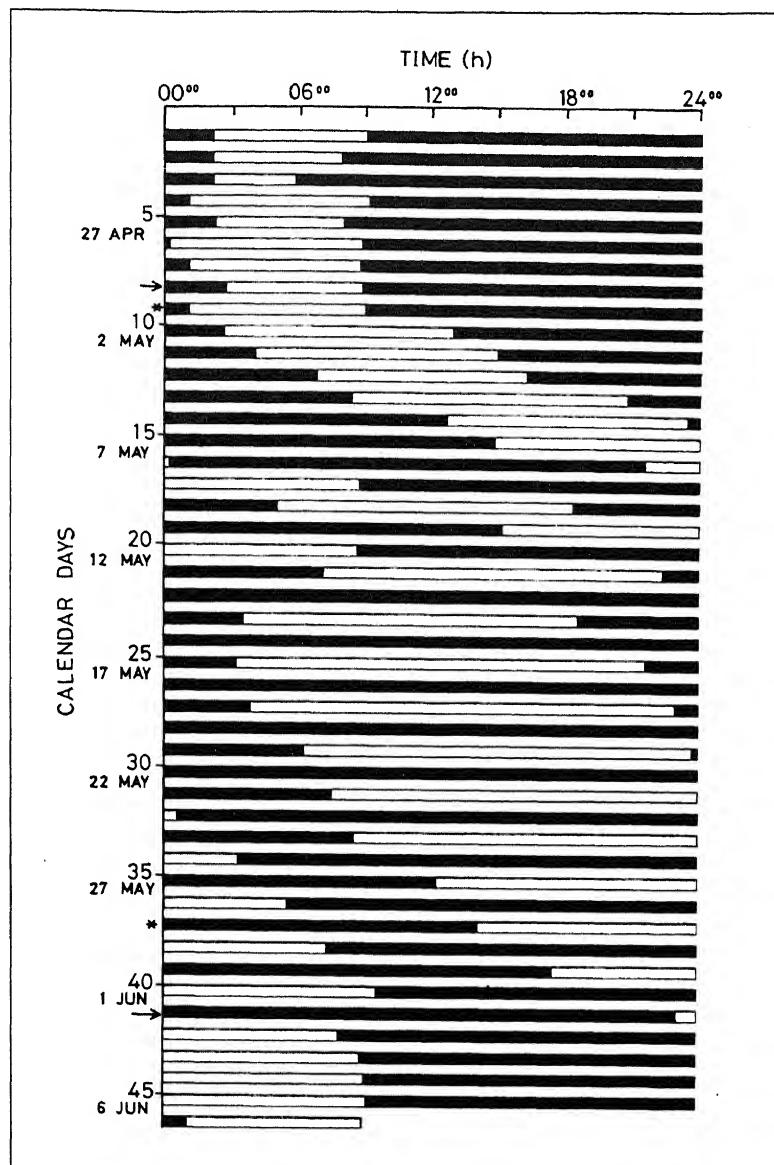
Figure 3 *The times of sleep and wakefulness of my second stay in isolation, plotted on a 24 hour scale. Each line in the graph represents one calendar day. Darkened bars indicate the period of wakefulness and hollow bars indicate period of sleep. The arrow on the 8th day (30.04.1991) represents the day I entered the isolation chamber and the one on 02.06.1991 represents the day I came out of it. The asterisk on 01.05.1991 indicates the day of onset of the first menstrual cycle and the one on 29.05.1991 indicates the day of onset of the second menstrual cycle. Note that the period of wakefulness and sleep increased every day and gradually stabilised towards the end of the stay. The maximum time I kept awake was around 34 hours and the maximum time that I slept was around 19 hours at a stretch. You can notice that some days I never slept at all! (for example 18th May, 20th May and 22nd May).*

per subjective day and the temperature had a period[5] of about 25.1 hours. In other words, my sleep-wake temperature and rhythm were in a state of 'desynchronisation'.

Generally, under normal societal conditions, both the sleep-wake cycle and the temperature cycle have the same period of about 24 hours. In such a condition, they are said to be 'synchronised'.

[5] Time interval between recurrences of a defined phase of the rhythm.

Figure 4 *The times of going to sleep and wakefulness of my second isolation plotted according to my subjective days. This is a multiple plot, plotted in such a way that it is easier to observe the course of the rhythm when you look at it. The first 8 days were pre-isolation data and each line here represents a calendar day. The period of isolation has been represented in subjective days (the days that I counted as opposed to the actual calendar days). Subjective day 1 starts on the 9th day in the graph and subjective day 20 is the last day of my isolation. Darkened bars indicate periods of wakefulness and hollow bars represent durations of sleep. Note that under pre and post isolation conditions the rhythm was in an entrained state, having an almost 24 hour rhythm and under isolation it free-runs. The period increased gradually and towards the end of the isolation period, it came to about 48 hours (circa bidian). Asterisks on the two occasions indicate onset of menstrual cycles. Note how the rhythm stabilises in the post-isolation period.*
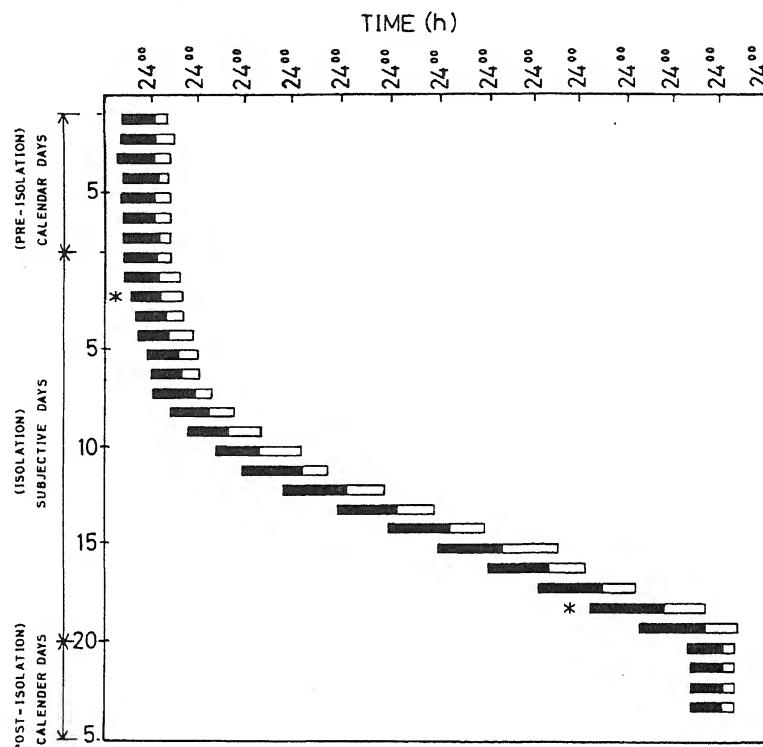
However, when a person is under isolation, the sleep-wake cycle may free-run with periods deviating from 24 hours. In such cases, the temperature cycle need not deviate, but can and most often will maintain its approximately 24 hour rhythmicity. Such a condition is called 'internal desynchronisation'. This is a proof that our bodily rhythms may be controlled by more than one clock.

Now, we had an interesting question to ask: would the menstrual cycle follow the sleep-wake cycle or the temperature cycle? Since I had experienced only 22 days, if it had to follow my sleep-wake cycle, it should occur after I experience 28 subjective days. On the other hand, if it follows the temperature cycle, it should have occurred after 28 calendar days. In my case, menstrual cycle occurred after 28 *calendar* days. It did not depend on the sleep-wake rhythm. Interestingly, re-adjustment to the societal condition occurred almost instantly. My sleep-wake rhythm started to have about a 24 hour period almost from the second day of my exit

from the chamber. You can imagine how strong the effect of social cues is to the human circadian rhythms! Since it is very difficult to get a female subject who has a circa-bidian rhythm, whose rhythm desynchronises and who has a regular 28 day menstrual cycle, it is difficult to repeat such experiments. Nevertheless, we wanted to test the validity of our results. The only way was to repeat the same experiment on the same subject i.e me! To exclude the influences of circannual rhythms (rhythms which repeat themselves in about a year) in the physiological parameters and seasonal variations if any, it was also decided to perform the experiment at the same time of the year, but two years later.

So I Did It Again!

So, for the second time in my life (in May 1991) I had another sojourn in the isolation chamber. Everything else was the same except that two years had elapsed and now I had more responsible activities to do, like reading reprints, analysing my data on the circadian rhythms of mice and so on. So, this time I spent my time in a more constructive fashion. The only precaution that I took was to take care that the knowledge of my previous results did not influence my thinking. This time my stay was truly terminated prematurely (after 32 calendar days of isolation) and there was a good reason behind that. When I was in the middle of my stay, Rajiv Gandhi was assassinated and the situation was chaotic outside. I think I must have been one of the very few people in the whole world who did not hear about it in time! Apart from the difficulties faced by the people outside in providing my require- ments, there was another problem associated with this ignorance. Every day I was supplied with my subjective day's newspaper. Since I was again having a circa-bidian rhythm and as a result, losing days, the monitoring committee could still keep providing me with newspapers without the news of assassination for some time. On 2nd of June, 1991, it was 22nd of May for me inside and they were not sure if I would endure the shock alone and hence asked me to come out on that day. So, my exit from the chamber the second time, lacked all the exhilarations of the first time

> The menstrual cycle occurred after 28 *calendar* days. It did not depend on the sleep-wake rhythm.

because of this news. It was a great shock to me indeed to hear the news. After coming out of the chamber, I had to continue wearing the solicorder for my post-isolation data. This time, I faced a peculiar problem. The solicorder unit has two wires attached to a small box and the box with the wires is quite conspicuous when one wears it. Just a few days after the assassination, imagine what would happen if a girl walked around with a box having two protruding wires!

The results of the second experiment proved the validity of the first one in the sense that we had obtained almost the same data as in the first experiment. This time my sleep-wake cycle had a period of about 46.1 hours and my temperature cycle, 24.4 hours. Desynchronisation occurred on subjective day 9 and menstrual cycle occurred exactly 28 calendar days later, thus confirming the results of the previous experiment. Re-entrainment to the societal conditions again occurred as earlier.

> The results of the second experiment proved the validity of the first one in the sense that we had obtained almost the same data as in the first experiment.

In addition, results of these two and other experiments that were performed earlier proved that there is a direct correlation between the 2 hour time estimation and the sleep-wakefulness rhythm. The subjects who had about a 24 hour rhythm estimated 2 hours almost correctly, whereas, subjects like me, who had a 48 hour rhythm estimated almost 6 hours as 2 hours. Hence, it is possible for the people outside to guess approximately how long the person is going to stay up that day, by observing the first 2 hour estimation of the day!

Where Can It Lead To?

One might pause to wonder why at all these experiments are being performed. Such studies can provide useful suggestions to people whose rhythm is disrupted such as shift workers, astronauts and people who travel across time-zones. Their sleep-wake rhythms measured under isolation and the period thus estimated, might provide information as to how efficient they would be while under disrupted day-night conditions. Their work load can be so sched-

uled as to match this efficiency period. For people who face problems of 'jet-lag' after inter-continental travel, such studies can furnish useful information about readjustment.

Results of these two and other experiments that were performed earlier proved that there is a direct correlation between the 2 hour time estimation and the sleep-wakefulness rhythm.

Suggested Reading

J Aschoff. Human Circadian Rhythms: a Multi-oscillator System. Federation Proceedings. 35, 2326-32. 1976.

J Aschoff. Circadian Rhythms in Man. in: Biological Time Keeping. (Ed) J Brady. Society for Experimental Biology Seminar Series 14. Cambridge University Press. 1982.

M K Chandrashekaran, L Geetha, G Marimuthu, R Subbaraj, P Kumarasamy, M S Ramkumar. The Menstrual Cycle in a Human Female under Social and Temporal Isolation is Not Coupled to the Circadian Rhythm in Sleep-Wakefulness. *Curr. Sci.* 60, (12) 703-705. 1991.

M K Chandrashekaran, G Marimuthu, R Subbaraj, P Kumarasamy, M S Ramkumar, K Sripathi. Direct Correlation Between the Circadian Sleep-Wakefulness Rhythm and Time Estimation in Humans under Social and Temporal Isolation. *J Biosci.* 16, 3, 97-101. 1991.

M K Chandrashekaran. Circadian Rhythms, Menstrual Cycles and Time Sense in Humans Under Social Isolation. in: Evolution of Circadian Clock. (Eds) T Hiroshige, K Honma. Hokkaido Univ Press Sapporo. 263-274. 1994.

R T W L Conroy, J N Mills. Human Circadian Rhythms. K & A Churchill, London. 1970.

M Siffre. Six Months Alone in a Cave. *National Geographic,* March 1975.

R Wever. The Circadian System of Man. Springer Verlag Berlin, Heidelberg, New York. 1979.

Address for correspondence
L Geetha
Centre for Ecological Sciences
Indian Institute of Science
Bangalore 560 012, India.
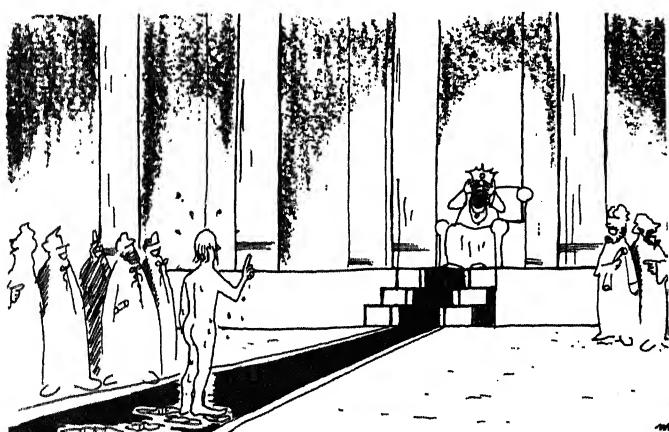
THAT'S FINE, ARCHIMEDES ! . . . BUT THIS IS NOT THE WAY TO GET MY ATTENTION!

MOHAN DEVADAS

On the Infinitude of the Prime Numbers

Euler's Proof

Shailesh A Shirali

Euclid's elegant proof that there are infinitely many prime numbers is well known. Euler proved the same result, in fact a stronger one, by *analytical* methods. This article gives an exposition of Euler's proof introducing the necessary concepts along the way.

Shailesh Shirali has been at the Rishi Valley School (Krishnamurti Foundation of India), Rishi Valley, Andhra Pradesh, for more than ten years and is currently the Principal. He has been involved in the Mathematical Olympiad Programme since 1988. He has a deep interest in talking and writing about mathematics, particularly about its historical aspects. He is also interested in problem solving (particularly in the fields of elementary number theory, geometry and combinatorics).

Introduction

In this article, we present Euler's very beautiful proof that there are infinitely many prime numbers. In an earlier era, Euclid had proved this result in a simple yet elegant manner. His idea is easy to describe. Denoting the prime numbers by $p_1, p_2, p_3, \ldots$, so that $p_1 = 2, p_2 = 3, p_3 = 5, \ldots$, he supposes that there are n primes in all, the largest being p_n. He then considers the number N where

$$N = p_1 p_2 p_3 \cdots p_n + 1,$$

and asks what the prime factors of N could be. It is clear that N is indivisible by each of the primes $p_1, p_2, p_3, \ldots, p_n$ (indeed, $N \equiv 1 \pmod{p_i}$ for each i, $1 \le i \le n$). Since every integer greater than 1 has a prime factorization, this forces into existence prime numbers other than the p_i. Thus there can be no largest prime number, and so the number of primes is infinite.

The underlying idea of Euler's proof is very different from that of Euclid's proof. In essence, he proves that the *sum of the reciprocals of the primes is infinite*; that is,

$$\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} + \cdots = \infty.$$

In technical language, the series $\Sigma_i\, 1/p_i$ *diverges*. Obviously, this cannot possibly happen if there are only finitely many prime numbers. The infinitude of the primes thus follows as a corollary. Note that Euler's result is stronger than Euclid's.

Convergence and Divergence

A few words are necessary to explain the concepts of convergence and divergence of infinite series. A series $a_1 + a_2 + a_3 + \cdots$ is said to *converge* if the sequence of partial sums,

$$a_1, a_1 + a_2, a_1 + a_2 + a_3, \cdots,$$

approaches some limiting value, say L; we write, in this case, $\sum_1^\infty a_i = L$. If, instead, the sequence of partial sums grows without any bound, we say that the series *diverges*, and we write, in short, $\sum_1^\infty a_i = \infty$.

Examples:

- The series $1/1 + 1/2 + 1/4 + \cdots + 1/2^n + \cdots$ converges (the sum is 2, as is easily shown).
- The series $1/1 + 1/3 + 1/9 + \cdots + 1/3^n + \cdots$ converges (the sum in this case is 3/2).
- The series $1+1+1+\cdots$ diverges (rather trivially).
- The series $1-1+1-1+1-1+1-\cdots$ also fails to converge, because the partial sums assume the values $1, 0, 1, 0, 1, 0, \ldots$, and this sequence clearly does not possess a limit.
- A more interesting example: $1 - 1/2 + 1/3 - 1/4 + \ldots$; a careful analysis shows that it too is convergent, the limiting sum being $\ln 2$ (the natural logarithm of 2)

Divergence of the Harmonic Series $\Sigma\, 1/i$

In order to prove Euler's result, namely, the divergence of $\Sigma\, 1/p_i$, we need to establish various subsidiary results. Along the way we

Euler proved that the sum of the reciprocals of the primes is infinite; the infinitude of the primes thus follows as a corollary.

A statement of the form $\Sigma\, a_i = \infty$ is to be regarded as merely a short form for the statement that the sums a_1, $a_1 + a_2$, $a_1 + a_2 + a_3$, ..., do not possess any limit. It is important to note that ∞ is *not* to be regarded as a number! We shall however frequently use phrases of the type '$x = \infty$' (for various quantities x) during the course of this article. The meaning should be clear from the context.

shall meet other examples of divergent series. To start with, we present the proof of the statement that

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} \cdots = \infty.$$

This rather non-obvious result is usually referred to as *the divergence of the harmonic series*. The proof given below is due to the Frenchman Nicolo Oresme and it dates to about 1350. We note the following sequence of equalities and inequalities:

$$\frac{1}{1} = \frac{1}{1},$$

$$\frac{1}{2} = \frac{1}{2},$$

$$\frac{1}{3} + \frac{1}{4} > \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

$$\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2},$$

$$\frac{1}{9} + \frac{1}{10} + \cdots + \frac{1}{16} > \frac{1}{16} + \frac{1}{16} + \cdots + \frac{1}{16} = \frac{1}{2},$$

> The earliest known proof of the divergence of the harmonic series is due to the Frenchman Nicolo Oresme and it dates to about 1350.

and so on. This shows that it is possible to group consecutive sets of terms of the series $1/1 + 1/2 + 1/3 + \cdots$ in such a manner that each group has a sum exceeding 1/2. Since the number of such groups is infinite, it follows that the sum of the whole series is itself infinite. (Note the crisp and decisive nature of the proof!)

Based on this proof, we make a more precise statement. Let $S(n)$ denote the sum

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n},$$

e.g., $S(3) = 11/6$. Generalizing from the reasoning just used, we find that

$$S(2^n) > 1 + \frac{n}{2}. \qquad (3.1)$$

(Please fill in the details of the proof on your own.) This means that by choosing n to be large enough, the value of $S(2^n)$ can be made to exceed any given bound. For instance, if we wanted the sum to exceed 100, then (3.1) assures us that a mere 2^{198} terms would suffice! This suggests the extreme slowness of growth of $S(n)$ with n. Nevertheless it does grow without bound; loosely stated, $S(\infty) = \infty$.

The result obtained above, (3.1), can also be written in the form,

$$S(n) > 1 + \frac{1}{2}\log_2 n.$$

> *Exercise*: Write out a proof of the above inequality.

A much more accurate statement can be made, but it involves calculus. We consider the curve Ω whose equation is $y = 1/x$, $x > 0$. The area of the region enclosed by Ω, the x-axis and the ordinates $x = 1$ and $x = n$ is equal to $\int_1^n \frac{1}{x}\,dx$, which simplifies to $\ln n$. Now let the region be divided into $(n - 1)$ strips of unit width by the lines $x = 1, x = 2, x = 3, \ldots, x = n$ (see *Figure 1*).

The growth of $S(n)$ with n is extremely slow. For the sum to exceed 100 we would require 2^{198} terms!

y = 1/x
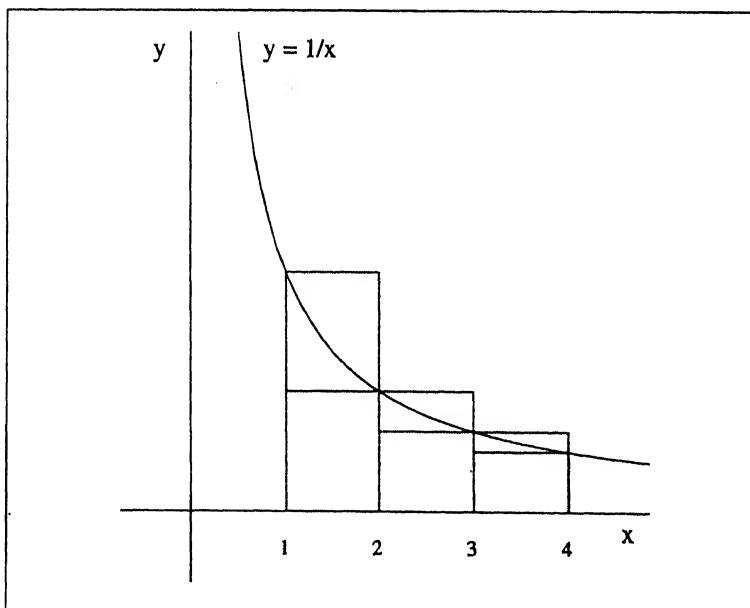
Figure 1 *The figure shows how to bound ln n by observing that ln n is the area enclosed by the curve y = 1/x, the x-axis and the ordinates x = 1 and x = n.*

Consider the region enclosed by Ω, the x-axis, and the lines $x = i - 1$, $x = i$. The area of this region lies between $1/i$ and $1/(i-1)$, because it can be enclosed between two rectangles of dimensions $1 \times 1/i$ and $1 \times 1/(i-1)$, respectively. (A quick examination of the graph will show why this is true.) By letting i take the values $2, 3, 4, \ldots, n$, and adding the inequalities thus obtained, we find that

$$\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} < \ln n < \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n-1}. \quad (3.2)$$

Relation (3.2) implies that

$$\ln n + \frac{1}{n} < \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} < \ln n + 1 \quad (3.3)$$

and this means that we have an estimate for $S(n)$ (namely, $\ln n + 0.5$) that differs from the actual value by no more than 0.5. A still deeper analysis shows that for large values of n, an excellent approximation for $S(n)$ is $\ln n + 0.577$, but we shall not prove this result here. It is instructive, however, to check the accuracy of this estimate. Write $f(n)$ for $\ln n + 0.577$. We now find the following:

$n =$	10	100	1000	10000	100000
$S(n) =$	2.92897	5.18738	7.48547	9.78761	12.0902
$f(n) =$	2.87959	5.18217	7.48476	9.78734	12.0899

The closeness of the values of $f(n)$ and $S(n)$ for large values of n is striking. (The constant 0.577 is related to what is known as the Euler-Mascheroni constant.)

In general, when mathematicians find that a series Σa_i diverges, they are also curious to know how *fast* it diverges. That is, they wish to find a function, say $f(n)$, such that the ratio $\left(\sum_1^n a_i\right) / f(n)$ tends to 1 as $n \to \infty$. For the harmonic series $\Sigma 1/i$, we see that one such function is given by $f(n) = \ln n$. This is usually expressed by saying that the harmonic series diverges like the logarithmic function. We note in passing that this is a very

In general, when mathematicians find that a series Σa_i diverges, they are also curious to know how *fast* it diverges.

slow rate of divergence, because $\ln n$ diverges more slowly than n^ε for any $\varepsilon > 0$, *no matter how small ε is*, in the sense that $\ln n / n^\varepsilon \to 0$ as $n \to \infty$ for every $\varepsilon > 0$. Obviously the function $\ln \ln n$ diverges still more slowly.

Exercise: Prove that if $a > 1$, then the series

$$\frac{1}{1^a} + \frac{1}{2^a} + \frac{1}{3^a} + \cdots$$

converges. (The conclusion holds no matter how close a is to 1, but it does not hold for $a = 1$ or $a < 1$, a curious state of affairs!) Further, use the methods of integral calculus (and the fact that for $a \neq 1$, the integral of $1/x^a$ is $x^{(1-a)}/(1-a)$ to show that the sum of the series lies between $1/(a-1)$ and $a/(a-1)$.

---

he fact that the sum $1/1 + 1/2^2 + 1/3^2 + \cdots$ is finite can be 10wn in another manner that is both elegant and elementary. ᵉe start with the inequalities, $2^2 > 1 \times 2, 3^2 > 2 \times 3, 4^2 > 3 \times 4$, ·, and deduce from these that

$$+ \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots < 1 + \frac{1}{1 \times 2} + \frac{1}{2 \times 3} + \frac{1}{3 \times 4} + \cdots$$

ıe sum on the right side can be written in the form,

$$1 + \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots, \qquad (3.4)$$

.ich (after a whole feast of cancellations) simplifies to $1 + 1/1$, ıt is, to 2. (This is sometimes described by stating that the series escopes' to 2.) Therefore the sum $1 + 1/2^2 + 1/3^2 + 1/4^2 + \cdots$ ᵉss than 2. We now call upon a theorem of analysis which states t if the partial sums of any series form an increasing sequence l are at the same time bounded, that is, they do not exceed some :d number, then they possess a limit. We conclude, therefore, t the series $\Sigma \, 1/i^2$ does possess a finite sum which lies between ıd 2.

The fact that the sum $1/1^2 + 1/2^2 + 1/3^2 + \ldots$ is finite can be shown in a manner that is both elegant and elementary.

The divergence of the harmonic series was independently proved by Johann Bernoulli in 1689 in a completely different manner. His proof is worthy of deep study, as it shows the counter-intuitive nature of infinity.

Bernoulli starts by assuming that the series $1/2 + 1/3 + 1/4 + \cdots$ (note that he starts with $1/2$ rather $1/1$) does have a finite sum, which he calls S. He now proceeds to derive a contradiction in the following manner. He rewrites each term occurring in S thus:

$$\frac{1}{3} = \frac{2}{6} = \frac{1}{6} + \frac{1}{6}, \quad \frac{1}{4} = \frac{3}{12} = \frac{1}{12} + \frac{1}{12} + \frac{1}{12}, \cdots,$$

and more generally,

$$\frac{1}{n} = \frac{n-1}{n(n-1)} = \frac{1}{n(n-1)} + \frac{1}{n(n-1)} + \cdots + \frac{1}{n(n-1)},$$

with $(n-1)$ fractions on the right side. Next he writes the resulting fractions in an array as shown below:

1/2	1/6	1/12	1/20	1/30	1/42	1/56	...
	1/6	1/12	1/20	1/30	1/42	1/56	...
		1/12	1/20	1/30	1/42	1/56	...
			1/20	1/30	1/42	1/56	...
				1/30	1/42	1/56	...
					1/42	1/56	...
						1/56	...

Note that the column sums are just the fractions $1/2, 1/3, 1/4, 1/5, \ldots$; thus S is the sum of all the fractions occurring in the array. Bernoulli now sums the rows using the telescoping technique used above (see equation (3.4)). Assigning symbols to the row sums as shown below,

$$A = \frac{1}{2} + \frac{1}{6} + \frac{1}{12} + \frac{1}{20} + \frac{1}{30} + \frac{1}{42} + \frac{1}{56} + \cdots,$$

$$B = \frac{1}{6} + \frac{1}{12} + \frac{1}{20} + \frac{1}{30} + \frac{1}{42} + \frac{1}{56} + \cdots,$$

$$C = \frac{1}{12} + \frac{1}{20} + \frac{1}{30} + \frac{1}{42} + \frac{1}{56} + \cdots,$$

$$D = \frac{1}{20} + \frac{1}{30} + \frac{1}{42} + \frac{1}{56} + \cdots,$$

he finds that:

$$A = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \left(\frac{1}{4} - \frac{1}{5}\right) + \cdots$$

$$= 1,$$

$$B = \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \left(\frac{1}{4} - \frac{1}{5}\right) + \left(\frac{1}{5} - \frac{1}{6}\right) + \cdots$$

$$= \frac{1}{2},$$

$$C = \frac{1}{3}, \qquad \text{(arguing likewise)},$$

$$D = \frac{1}{4},$$

Bernoulli's proof is worthy of deep study, as it shows the counter-intuitive nature of infinity.

and so on. Thus the sum S, which we had written in the form $A + B + C + D + \cdots$, turns out to be equal to

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots \quad .$$

Now this looks disappointing -- just as things were beginning to look promising! We seem to have just recovered the original series after a series of very complicated steps. But in fact something significant has happened: *an extra '1' has entered the series*. At the start we had defined S to be $1/2 + 1/3 + 1/4 + \cdots$; now we find that S equals $1 + 1/2 + 1/3 + 1/4 + \cdots$. This means that $S = S + 1$. However, no finite number can satisfy such an equation. Conclusion: $S = \infty$!

There are many other proofs of this beautiful result, but I shall leave you with the pleasant task of coming up with them on your own. Along the way you could set yourself the task of proving that each of the following sums diverge:

- $1/1 + 1/3 + 1/5 + 1/7 + 1/9 + \cdots$;
- $1/1 + 1/11 + 1/21 + 1/31 + 1/41 + \cdots$;
- $1/a + 1/b + 1/c + 1/d + \cdots$, where $a, b, c, d, \ldots$, are the successive terms of any increasing arithmetic progression of positive real numbers.

Elementary Results

The next result that we shall need is the so-called fundamental theorem of arithmetic: *every positive integer greater than 1 can be expressed in precisely one way as a product of prime numbers.* We shall not prove this very basic theorem of number theory. For a proof, please refer to any of the well-known texts on number theory, e.g., the text by Hardy and Wright, or the one by Niven and Zuckermann.

We shall also need the following rather elementary results: (i) if k is any integer greater than 1, then

$$\frac{1}{1 - 1/k} = 1 + \frac{1}{k} + \frac{1}{k^2} + \frac{1}{k^3} + \frac{1}{k^4} + \cdots ,$$

which follows by summing the geometric series on the right side, and (ii) if a_i, b_j are any quantities, then

$$\left(\sum_i a_i \right) \left(\sum_j b_j \right) = \sum_{i,j} a_i b_j ,$$

where, in the sum on the right, each pair of indices (i, j) occurs *precisely once*.

Now consider the following two equalities, which are obtained from (4.1) using the values $k = 2, k = 3$:

> The fundamental theorem of arithmetic states that every positive integer greater than one can be expressed in precisely one way as a product of prime numbers.

$$\frac{1}{1 - \frac{1}{2}} = 1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \cdots,$$

$$\frac{1}{1 - \frac{1}{3}} = 1 + \frac{1}{3} + \frac{1}{3^2} + \frac{1}{3^3} + \frac{1}{3^4} + \cdots.$$

We multiply together the corresponding sides of these two equations. On the left side we obtain $2 \times 3/2 = 3$. On the right side we obtain the product

$$(1 + 1/2 + 1/2^2 + 1/2^3 + \cdots) \times (1 + 1/3 + 1/3^2 + 1/3^3 + \cdots)$$

Expanding the product, we obtain:

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \cdots$$

$$+ \frac{1}{6} + \frac{1}{12} + \frac{1}{24} + \cdots + \frac{1}{18} + \frac{1}{36} + \frac{1}{72} + \cdots,$$

that is, we obtain the sum of the reciprocals of all the positive integers that have only 2 and 3 among their prime factors. The fundamental theorem of arithmetic assures us that each such integer occurs *precisely once* in the sum on the right side. Thus we obtain a nice corollary: if A denotes the set of integers of the form $2^a 3^b$, where a and b are non-negative integers, then

$$\sum_{z \in A} \frac{1}{z} = 3.$$

We obtain a nice corollary: if A denotes the set of integers of the form $2^a 3^b$, where a and b are non-negative integers, then

$$\sum_{z \in A} 1/z = 3.$$

If we multiply the left side of this relation by $(1 + 1/5 + 1/5^2 + 1/5^3 + \cdots)$ and the right side by $3/(1 - \frac{1}{5})$, we obtain the following result:

$$\sum_{z \in B} \frac{1}{z} = \frac{3}{1 - \frac{1}{5}} = \frac{15}{4},$$

where B denotes the set of integers of the form $2^a\,3^b\,5^c$, where a, b and c denote non-negative integers.

Continuing this line of argument, we see that infinitely many such statements can be made, for example:

- If C denotes the set of positive integers of the form $2^a\,3^b\,5^c\,7^d$, where a, b, c and d are non-negative integers, we then have $\sum_{z\in C} 1/z = (15/4)\,(7/6) = 35/8$.
- If D denotes the set of positive integers of the form $2^a\,3^b\,5^c\,7^d\,11^e$, then $\sum_{z\in D} 1/z = (35/8)\,(11/10) = 77/16$.

Infinitude of the Primes

Suppose now that there are only finitely many primes, say $p_1, p_2, p_3, \ldots, p_n$, where $p_1 = 2, p_2 = 3, p_3 = 5, \ldots$. We consider the product

$$\frac{1}{1 - \tfrac{1}{2}} \; \frac{1}{1 - \tfrac{1}{3}} \; \frac{1}{1 - \tfrac{1}{5}} \; \cdots \; \frac{1}{1 - \tfrac{1}{p_n}}$$

This is obviously a finite number, being the product of finitely many non-zero fractions. Now this product also equals

$$\left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots\right) \times \left(1 + \frac{1}{3} + \frac{1}{3^2} + \cdots\right) \times$$

$$\left(1 + \frac{1}{5} + \frac{1}{5^2} + \cdots\right) \times \; \cdots \; \times \left(1 + \frac{1}{p_n} + \frac{1}{p_n^2} + \cdots\right)$$

When we expand out this product, we find, by continuing the line of argument developed above, that we obtain *the sum of the reciprocals of all the positive integers.* To see why, we need to use the fundamental theorem of arithmetic and the assumption that 2, 3, 5, ..., p_n are *all* the primes that exist; these two statements together imply that every positive integer can be expressed *uniquely* as a product of non-negative powers of the n primes 2, 3, 5, ... p_n. From

this it follows that the expression on the right side is precisely the sum

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots,$$

written in some permuted order. But by the Oresme-Bernoulli theorem, the latter sum is infinite! So we have a contradiction: the finite number

$$\frac{1}{1 - \frac{1}{2}} \quad \frac{1}{1 - \frac{1}{3}} \quad \frac{1}{1 - \frac{1}{5}} \quad \cdots \quad \frac{1}{1 - \frac{1}{p_n}}$$

has been shown to be infinite -- an absurdity! The only way out of this contradiction is to drop the assumption that there are only finitely many prime numbers. Thus we reach the desired objective, namely, that of proving that there are infinitely many prime numbers.

Note that, as a bonus, there are several formulas that drop out of this analysis, more or less as corollaries. For instance, we find that

$$\frac{1}{1 - \frac{1}{2^2}} \quad \frac{1}{1 - \frac{1}{3^2}} \quad \frac{1}{1 - \frac{1}{5^2}} \quad \cdots = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots$$

that is, the infinite product and the infinite sum both converge to the same (finite) value. By a stunning piece of reasoning, including a few daring leaps that would leave today's mathematicians gasping for breath, Euler showed that both sides of the above equation are equal to $\pi^2/6$. Likewise, we find that

$$\frac{1}{1 - \frac{1}{2^4}} \quad \frac{1}{1 - \frac{1}{3^4}} \quad \frac{1}{1 - \frac{1}{5^4}} \quad \cdots = 1 + \frac{1}{2^4} + \frac{1}{3^4} + \frac{1}{4^4} + \cdots,$$

and this time both sides converge to $\pi^4/90$. Euler proved all this and much much more; it is not for nothing that he is at times referred to as *analysis incarnate!*

It is not for nothing that Euler is at times referred to as analysis incarnate!

Leonhard Euler

Since universities were not the major research centres in his days, Leonhard Euler (1707-1783) spent most of his life with the Berlin and Petersburg Academies. Pious, but not dogmatic, Euler conducted prayers for his large household, and created mathematics with a baby on his lap and children playing all around. Euler withheld his own work on calculus of variations so that young Lagrange (1736-1813) could publish it first, and showed similar generosity on many other occasions. Utterly free of false pride, Euler always explained how he was led to his results saying that "the path I followed will perhaps be of some help". And, indeed, generations of mathematicians followed Laplace's advice: "Read Euler, he is our master in all!".

The Divergence of $\Sigma\, 1/p$

As mentioned earlier, Euler showed in addition that the sum

$$\sum_{i \geq 1} \frac{1}{p_i} = \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \cdots$$

is itself infinite. We are now in a position to obtain this beautiful result. For any positive integer $n \geq 2$, let P_n denote the set of prime numbers less than or equal to n. We start by showing that

$$\prod_{p \in P_n} \frac{1}{1 - 1/p} > \sum_{j=1}^{n} \frac{1}{j}. \tag{6.1}$$

Our strategy will be a familiar one. We write down the following inequality for each $p \in P_n$, which follows from equation (4.1):

$$\frac{1}{1 - 1/p} > 1 + \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \cdots + \frac{1}{p^n}.$$

The '>' sign holds because we have left out all the positive terms that follow the term $1/p^n$. Multiplying together the corresponding sides of all these inequalities ($p \in P_n$), we obtain:

$$\prod_{p \in P_n} \frac{1}{1 - 1/p} > \prod_{p \in P_n} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \cdots + \frac{1}{p^n} \right)$$

When we expand out the product on the right side, we obtain a sum of the form $\sum_{j \in A} 1/j$ for some set of positive integers A. This set certainly includes all the integers from 1 to n because the set P_n contains all the prime numbers between 1 and n. Inequality (6.1) thus follows immediately.

Next, we already know (see equation (3.3)) that

$$\sum_{j=1}^{n} \frac{1}{j} > \ln n + \frac{1}{n} > \ln n. \tag{6.2}$$

Combining (6.1) and (6.2), we obtain the following inequality:

$$\prod_{p \in P_n} \frac{1}{1 - 1/p} > \ln n.$$

Taking logarithms on both sides, this translates into the statement,

$$\sum_{p \in P_n} \ln \left(\frac{1}{1 - 1/p} \right) > \ln \ln n. \tag{6.3}$$

Our task is nearly over. It only remains to relate the sum $\sum_{p \in P_n} 1/p$ with the sum on the left side of (6.3). We accomplish this by showing that the inequality

$$\frac{7x}{5} > \ln \frac{1}{1-x} \tag{6.4}$$

holds for $0 < x \le 1/2$.

To see why (6.4) is true, draw the graph of the curve Γ whose equation is $y = \ln (1/(1-x))$, over the domain $-\infty < x < 1$, (see *Figure 2*). Note that Γ passes through the origin and is convex over
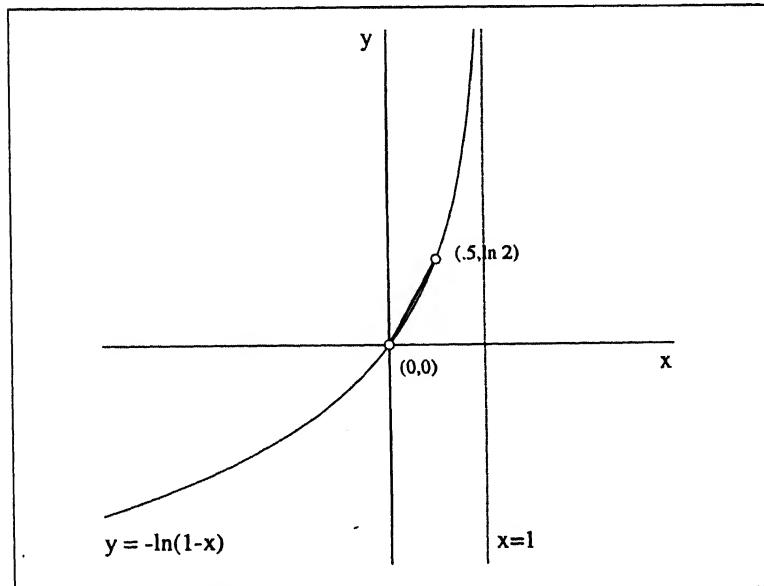
y

(.5, ln 2)

(0,0)

x

y = -ln(1-x)

x=1

Figure 2 The graph shows that for $0 \le x \le 1/2$, we have $(2 \ln 2)x \ge \ln (1/(1-x))$.

its entire extent. (Proof: Write $f(x) = -\ln(1-x)$; then $f'(x) = 1/(1-x)$ and $f''(x) = 1/(1-x)^2 > 0$ for all $x < 1$.)

The convexity of Γ implies that the chord joining the points $A(0,0)$ and $B(1/2, \ln 2)$ lies completely *above* the curve. The equation of AB is $y = (2 \ln 2)\, x$, so over the range $0 \le x \le 1/2$ we have the inequality:

$$(2 \ln 2)\, x \ge \ln\left(\frac{1}{1-x}\right).$$

Since $\ln 2 \approx 0.69315 < 0.7 = 7/10$, (6.4) follows.

Inequality (6.4) implies that

$$x > \frac{5}{7} \ln\left(\frac{1}{1-x}\right)$$

for $x = 1/2, x = 1/3, x = 1/5, \ldots$. Therefore, by addition,

$$\sum_{p \in P_n} \frac{1}{p} > \frac{5}{7}\left(\sum_{p \in P_n} \ln \frac{1}{1 - 1/p}\right). \tag{6.5}$$

Combining (6.3) and (6.5), we deduce that

$$\sum_{p \in P_n} \frac{1}{p} > \frac{5}{7} \ln \ln n.$$

As $n \to \infty$, the right side diverges to infinity, therefore so does the left side, so we reach our desired objective, that of showing the divergence of $\sum_i 1/p_i$.

An Alternate Proof

Here is an alternate proof of the claim that $\sum_i 1/p_i$ diverges. The proof has been written in an 'old-fashioned' style and purists will protest. Nevertheless, we shall present the proof and let readers

judge for themselves. Let S denote the sum $\Sigma_i \, 1/p_i$. We shall make use of the following result:

$$e^x \geq 1 + x \quad \text{for all real values of } x,$$

with equality holding precisely when $x = 0$. The graphs of e^x and $1+x$ show why this is true; the former graph is convex over its entire extent (examine the second derivative of e^x to see why), while the latter, a line, is tangent to the former at the point $(0,1)$, and lies entirely below it everywhere else. Substituting the values $x = 1/2, x = 1/3, x = 1/5, \ldots$, successively into this inequality, we find that

$$e^{1/2} > 1 + \frac{1}{2}, \quad e^{1/3} > 1 + \frac{1}{3}, \quad e^{1/5} > 1 + \frac{1}{5}, \quad \cdots .$$

Multiplying together the corresponding sides of these inequalities, we obtain:

$$e^S > \left(1 + \frac{1}{2}\right) \left(1 + \frac{1}{3}\right) \left(1 + \frac{1}{5}\right) \cdots .$$

The infinite product on the right side yields the following series:

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{10} + \frac{1}{11} + \frac{1}{13} + \frac{1}{14} + \frac{1}{15} + \cdots .$$

This series is the sum of the reciprocals of all the positive integers whose prime factors are all distinct; equivalently, the positive integers that have no squared factors. These numbers are sometimes referred to as the *quadratfrei* or *square-free* numbers. Let Q denote this sum. We shall show that this series itself diverges, in other words, that $Q = \infty$. This will immediately imply that $S = \infty$ (for $e^S > Q$), and Euler's result will then follow.

We consider the product

$$Q \times \left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots\right)$$

This product, when expanded out, gives the following series:

Johann Bernoulli

Johann Bernoulli (1667-1748), the younger brother of Jakob Bernoulli (1654-1705) took it upon himself to spread Leibniz's calculus across the European continent. Johann's proof of the divergence of the harmonic series first appeared (1689) in Jakob's treatise and with uncharacteristic fraternal affection, Jakob even prefaced the argument with an acknowledgement of his brother's priority.

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots,$$

that is, we obtain the harmonic series. To see why, note that every positive integer n can be *uniquely* written as a product of a square-free number and a square; for example, $1000 = 10 \times 10^2$, $2000 = 5 \times 20^2$, $1728 = 3 \times 24^2$, and so on. Now when we multiply

$$\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{10} + \frac{1}{11} + \frac{1}{13} + \frac{1}{14} + \frac{1}{15} + \cdots\right)$$

with

$$\left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots\right)$$

we find, by virtue of the remark just made, that the reciprocal of each positive integer n occurs *precisely once* in the expanded product. This explains why the product is just the harmonic series. Now recall that the sum

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots$$

is finite (indeed, we have shown that it is less than 2). It follows that

$$Q \times (\text{some finite number}) = \infty.$$

Therefore $Q = \infty$, and Euler's result ($\Sigma_i \, 1/p_i = \infty$) follows. QED!

Readers who are unhappy with this style of presentation, in which ∞ is treated as an ordinary real number, will find it an interesting (but routine) exercise to rewrite the proof to accord with more exacting standards of rigour and precision.

Conclusion

A much deeper – but also more difficult – analysis shows that the sum $1/p_1 + 1/p_2 + 1/p_3 + \ldots + 1/p_n$ is approximately equal to

Suggested Reading

G H Hardy, E M Wright. An Introduction to the Theory of Number. 4th ed., Oxford, Clarendon Press, 1960.

Ivan Niven, Herbert S Zuckermann. An Introduction to the Theory of Numbers. Wiley Eastern Ltd., 1989.

Tom Apostol. An Introduction to Analytic Number Theory. Narosa Publishing House. 1979.

XVI. *Summa seriei infinitæ harmonicè progreſſionalium ,* $\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}$ *&c. eſt infinita.*

Id primus deprehendit Frater : inventa namque per præced. ſumma ſeriei $\frac{1}{2} + \frac{1}{6} + \frac{1}{12} + \frac{1}{25} + \frac{1}{30}$, &c. viſurus porrò, quid emergeret ex iſta ſerie , $\frac{1}{2} + \frac{2}{6} + \frac{3}{12} + \frac{4}{20} + \frac{5}{30}$, &c. ſi reſolveretur methodo Prop. XIV. collegit p opoſitionis veritatem ex abſurditate manifeſta, quæ ſequeretur , ſi ſumma ſeriei harmonicæ finita ſtatueretur.   Animadvertit enim ,

Seriem A , $\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}$, &c. ∞ (fractionibus ſingulis in alias, quarum numeratores ſunt 1, 2 , 3 , 4 , &c. transmutatis)

ſeriei B , $\frac{1}{2} + \frac{1}{6} + \frac{3}{12} + \frac{4}{20} + \frac{5}{30} + \frac{6}{42}$, &c. ∞ C + D + E + F, &c.

C. $\frac{1}{2} + \frac{1}{6} + \frac{1}{12} + \frac{1}{25} + \frac{1}{30} + \frac{1}{42}$, &c. ∞ per præc. $\frac{1}{1}$ ⎤

D.. $+ \frac{1}{6} + \frac{1}{12} + \frac{1}{20} + \frac{1}{30} + \frac{1}{42}$, &c. ∞ C − $\frac{1}{2}$ ∞ $\frac{1}{2}$ ⎬ ∞ G ;

E. . . $+ \frac{1}{12} + \frac{1}{20} + \frac{1}{30} + \frac{1}{42}$, &c. ∞ D − $\frac{1}{6}$ ∞ $\frac{1}{3}$ ⎪ unde

F. . . . . . $+ \frac{1}{20} + \frac{1}{35} + \frac{1}{42}$, &c. ∞ E − $\frac{1}{12}$ ∞ $\frac{1}{4}$ ⎪ ſequi-

&c. ∞   &c. ⎦ tur, ſe-

(riem G ∞ A , totum parti, ſi ſumma finita eſſet.

Ego

ln ln n. This is usually stated in the following form: as n tends to ∞, the fraction

$$\frac{1/p_1 + 1/p_2 + 1/p_3 + \cdots + 1/p_n}{\ln \ln n}$$

tends to 1. This is indeed a striking result, reminiscent of the earlier result that $1/1 + 1/2 + 1/3 + \cdots + 1/n$ is approximately equal to ln n. It shows the staggeringly slow rate of divergence of the sum of the reciprocals of the primes. The harmonic series $\Sigma_i 1/i$ diverges slowly enough – to achieve a sum of over 100, for instance, we would need to add more than 10^{43} terms, so it is certainly not a job that one can leave to finish off over a weekend. (Do you see where the number 10^{43} comes from?) On the other hand, to achieve a sum of over 100 with the series $\Sigma_i 1/p_i$, we need to add something like $10^{10^{43}}$ terms!! This number is so stupendously large that it is a hopeless task to make any visual image of it. Certainly there is no magnitude even remotely comparable to it in the whole of the known universe.

Address for correspondence
Shailesh Shirali
Rishi Valley School
Rishi Valley 517 352,
Chittoor Dist. (A P), India.

What's New in Computers

Intel's New P6 Processor

Vijnan Shastri

This article briefly describes the architecture of a new microprocessor, called P6, being released by Intel Corporation.

The P6 is the next processor to be released (sometime in 1996) by Intel after the Pentium—currently the latest processor from them. P6 incorporates many advanced features such as superpipelining, super-scalar architecture, branch-prediction and advanced caching. We will explore these features (see page 99) in the following paragraphs and learn what these terminologies mean.

Super-Pipelining and Super-Scalar Architecture

Any processor is in fact a complex state machine. A state machine is a digital system driven by a clock that goes through various states (referred to as state transitions) on every 'tick' (clock cycle). The transitions through states depend on two sets of inputs: the current state and the outside world inputs. The outputs of the state machine are decoded from (obtained from) the states of the machine. For instance, if you were to build a digital stop watch, then you would build a state machine with a clock period of one second and the inputs would be the current time, start, stop and reset. The most important input for a processor is the set of instructions it fetches. The main steps involved in this process are *instruction fetch, decode* and *execute*. Each of these operations takes some time. In early processors such as 8085, these operations took place serially because the state machine was designed as a single monolithic unit. However, in current microprocessors, the state machine is divided into independent units (which interact with one another). This means that the instruction fetch unit fetches an instruction and hands it over to the decoding unit. While the decoding goes on, the instruction fetch unit fetches the next

Vijnan Shastri works at the Centre for Electronics Design Technology in the Indian Institute of Science, Bangalore, and his areas of interest are digital and microprocessor systems, storage subsystems and systems software.

A state machine is a digital system driven by a clock that goes through various states (referred to as state transitions) on every 'tick' (clock cycle).

instruction. When the current instruction is being executed by the execution unit, the decoding unit decodes the next instruction and so on. Note that this means that the processor executes one instruction every clock cycle. Hence, ideally all units are busy all the time and this process is referred to as pipelining. This is shown in *Figure 1*, where a simple pipeline of three units is illustrated.

We draw the following analogy: In a bread-factory assembly line, one person fetches the bread, the one next to him slices the bread with a slicing machine, the next one packs it, and the last one stores the packed bread in the shelf and maintains the count. Of course, both in this situation and in the processor, all units of the pipeline must take the same amount of time (typically one clock cycle) to do their bit. If this is not done, the units taking less time will remain partly idle in every clock cycle. This may be desirable (and humane!) in a human pipeline but will lead to performance degradation due to under-utilization of the processor. Hence, designers are forced to keep the decoding unit (which is usually the most time consuming process) simple. Keeping the decoding unit simple means having a reduced number of instructions. This is the idea behind the RISC (reduced instruction set computer) as opposed to CISC (complex instruction set computer). The 8085 and 8086 for instance, are CISC implementations.

We have spoken about three basic units: *fetch, decode* and *execute*.

The most important input for a processor is the instructions it fetches. The main steps involved in this process are instruction fetch, decode and execute.

The P6 contains fourteen such independent units (called stages), as compared to the Pentium's five stages. Hence it is said to be *super-pipelined*. In addition to pipelining, the independence of the units such as the integer units (two of them), floating-point unit and address generator units allows the P6 to simultaneously execute up to five instructions per clock cycle instead of one in the case of the 80486 and two in the case of the Pentium. This is why the P6 is called *super-scalar*. To further help this feature, the P6 supports out-of-order execution of the sequential instruction stream (i.e., re-orders them) such that the independent units are all kept busy. This is done without affecting the integrity of the program. In other words, the designers have done their best to build-in circuitry to try and keep all the units busy all the time. This does not always happen because programs often contain instructions that depend on the previous instruction. Instruction B for instance is said to be dependent on instruction A, if the execution of B depends on the result of the execution of A.

Branch Prediction

We have seen the pipe-lined structure of the P6. We also know that the processor fetches its instructions one-by-one from successive memory locations. This is known as the program flow. Often, a break occurs in the program flow. This means the processor must begin fetching instructions from some other location (rather than a successive location) and continue fetching from that location. These breaks are referred to as 'jumps' since the processor jumps to a new location to continue its operation. Jumps occur whenever there is a conditional instruction or when there are loops in a program. This is illustrated below:

1.   if number A is greater than number B
2.   then multiply A by number C
3.   else multiply number B by number C
4.   fetch number D.

Features of the P6 Processor

- Has a super-pipeline consisting of 14 stages.
- First version operates on a 133 Mhz clock and many instructions take a single clock cycle to execute.
- Super-scalar architecture capable of issuing upto 3 instructions at a time.
- 5 parallel execution units: 2 integer, one load, one store and one floating point unit (FPU).
- Supports out-of-order execution of instructions.
- One internal cache of 16 Kbytes (8 kbytes for code and 8 kbytes for data).
- A second cache (on a separate die but on the same package) of 256 kbytes with a separate 64-bit data path.
- P6 is about 800 times faster than a 8086 and about 2.5 times faster than a Pentium.
- Dissipates 20 watts of power.
- Estimated initial price : $1500.

Now, while executing step 1 if the processor finds that number A is indeed greater than number B it will fetch successive instructions of step 2 (no jump occurs). But after executing step 2 it will jump to step 4 and there is a break. If the processor finds at step 1 that number A is less than number B then it jumps to step 3, but no jump occurs between step 3 and step 4.

An example of a loop where the program calculates the power of 2 given the exponent 'n' is as follows:

```
1.    count=0, answer=1
2.    do steps 3 and 4 until count = n
3.        answer= 2*answer
4.        increment count
5.    store answer
```

Until the count is 'n' the program will jump back from step 4 to step 2 (where the value of count is checked).

Every time there is a break in program flow the overhead on the instruction fetch unit to go to a new location and start fetching again is quite high and this leads to 'bubbles' in the pipeline. The

Bubbles are formed in the pipeline whenever there is a break in program flow.

bubbles refer to the fact that some units will be idle for one or more clock cycles depending on the overhead. Let us go back to our bread-factory analogy. Notice that the person who fetches the bread would rather go and fetch many loaves of bread (from the oven) than only one. That is because of the overhead he incurs to go, pick and return — he might as well pick up several rather than one. This is exactly what happens in a processor. It is far more efficient to fetch more than one instruction at a time from memory and store them in an 'instruction queue' in readiness for the decoding unit. This is called 'pre-fetching'. Let us suppose that the factory produces two varieties of bread (milk bread and sweet bread) and the last person in the pipeline (who is also keeping count) tells the first one: " Enough of milk breads, we need to produce sweet breads now". The first person has go to another oven and fetch the sweet breads. Till that time the others will be idle. On the other hand if the last person can predict in advance and accordingly inform the first person to fetch sweet breads, then there won't be bubbles and no person will be idle. The branch prediction unit in the P6 does something similar. It uses statistical techniques to predict whether a (jump) branch will be taken or not and accordingly does the prefetching to avoid bubbles. Since this is probabilistic, there are times when the predictions are not right and there is a temporary dip in performance.

> The branch prediction unit in the P6 uses statistical techniques to predict whether a (jump) branch will be taken or not and accordingly does the pre-fetching to avoid bubbles.

Caching

The powerful computing engine capable of high speed data processing has to be supplied with data at the required rate. This is difficult to do if one considers the memory technology and the delays introduced by external electrical connections. To achieve this, they have built two caches for the chip; (caching was explained separately in an earlier article of the series). One of these (the level-one or L1 cache) is 16 kilobytes in size, and is integrated with the processor core on the same silicon die which consists of around 5.5 million transistors. The second cache (L2 cache) is 256 kilobytes in size and built on a separate silicon die consisting of 15.5 million transistors. The two dies are then
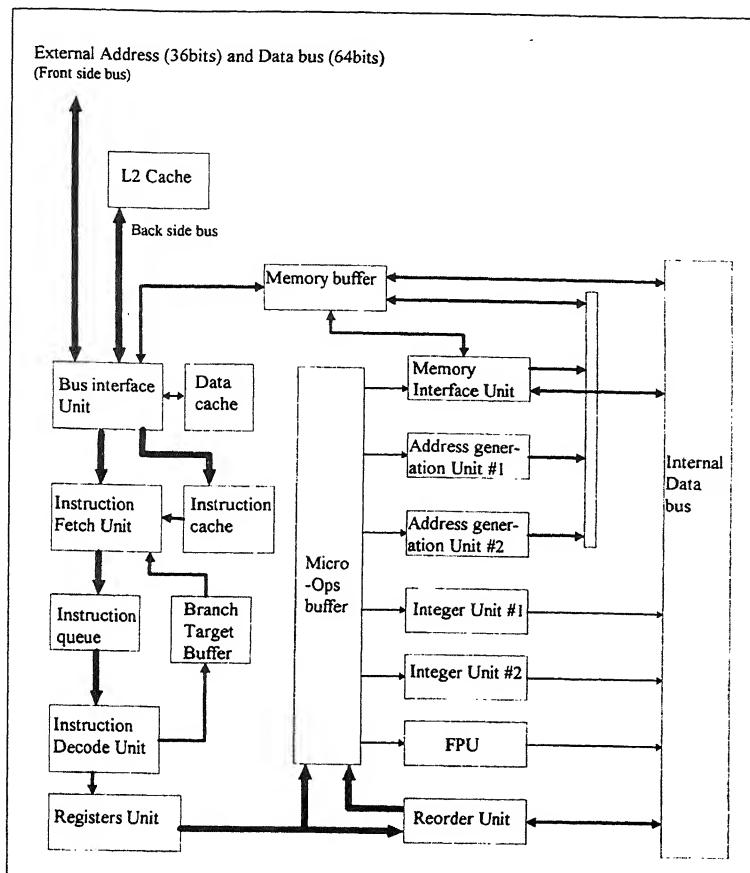
The P6 can run 16-bit software but will do so in an inefficient manner. In fact the performance of P6 will be slower than the Pentium for 16bit software!

integrated on the same ceramic package and bonded together by a technique called wire-bonding. This package, which has a total of 387 pins, is known as an MCM (multichip module). The L2 cache has its own path to the bus interface unit of the processor, separate from the path to the main memory. Both these data paths are 64 bits wide. This sophisticated design of caches ensures that the data is supplied to the P6 at a rate that matches the consumption rate of the processor. The existence of the L2 cache also saves system designers (designing boards with the P6 as the CPU) the trouble of building an external cache system and thus simplifies board design considerably. There is one caveat however to all this performance enhancement. The P6 designers expected that by the time it was released, users would be predominantly running 32-bit software and so they optimized the architecture for this kind

of software. The P6 can run 16-bit software but will do so in an inefficient manner. In fact, the performance of P6 will be slower than the Pentium for 16-bit software! This is the price that Intel has to pay to carry the 'DOS baggage' (see an earlier article in *Resonance* Vol.1, No.1, on Windows 95) in all its processors.

Conclusion

Although P6 is a high-performance microprocessor with a feature rich architecture, it has competition from other processors such as AMD's K5, NexGen's Nx586 and Cyrix's 5x86 which run x86 code and all of which share architectural innovations similar to the P6. The markets will ultimately decide which one succeeds. Meanwhile, Intel has started development of the P7, which will boost performance levels even further.

Address for correspondence
Vijnan Shastri
Centre for Electronics
Design Technology,
Indian Institute of Science,
Bangalore 560 012, India.

Suggested Reading

'P6 the Next Step?'. *PC Magazine*. September 12, 1995.
'Intel's P6'. *Byte* . April 1995.

NOW DOES THIS EXPLAIN THE CHOICE OF THE WALL PAPER, SIR ?
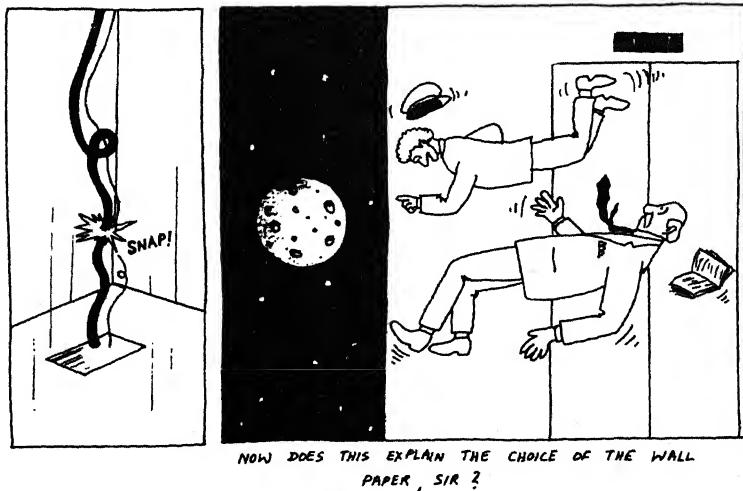
MOHAN DEVADAS

Nature Watch

Diversity of Bats

G Marimuthu

The author introduces us to the fascinating world of bats: their features, classification, habitats, food habits, and distribution.

Unique Characteristics of Bats

Bats are the only mammals that can fly. They are unique because of their capacity for flight and echolocation and their ability to hang upside down. Zoologists place the bats under the order known as Chiroptera. In Greek, the word *chiro* means hand, and *ptera* means wings. Their hands are modified to form a wing membrane which is a fold of skin stretched from the sides of the body to the elongated finger bones. The thumbs are free from the stretch of these wing membranes. The wings are divided into separate compartments by the elongated fingers. In this way, bats differ from the pterosaurs (the extinct flying reptile) whose wings were also folds of skin, but supported by a single elongated finger. The hind legs of bats also support the wing membranes. A few species of bats have a short or long tail which is either partly enclosed by the tail (interfemoral) membrane or extends between the two legs.

Bats are found in all parts of the world except the Arctic and Antarctic regions. The order Chiroptera comprising of nearly 850 species is the second largest in the world coming right after rodents (mice, squirrels, etc.). *Table 1* lists the 18 different families of bats, the number of species, their distribution and the type of food they prefer.

Classification

Bats are arranged into two major categories or suborders: Megachiroptera and Microchiroptera. As the name implies,

G Marimuthu, affectionately called "Batman" has made pioneering studies on the behaviour and ecology of Indian bats. How bats catch frogs has been a major theme of his research.

Bats are unique because of their capacity for flight and echolocation and their ability to hang upside down.

		Distribution	Diet	No. of species
***	Suborder			
**	Super family			
*	Family			

Table 1. Diversity of bats with their classification, distribution and diet.

		Distribution	Diet	No. of species
***	*Megachiroptera*			
*	Pteropodidae (Old world fruit bats, flying foxes)	Old World Tropics	fruits, nectar and flowers	150
***	*Microchiroptera*			
**	Emballonuroidea			
*	Rhinopomatidae (Mouse-tailed bats)	Africa, Asia and Borneo	insects	2
*	Craseonycteridae (Hog-nosed bats)	Thailand	insects	1
*	Emballonuridae (Sheath-tailed bats)	Tropics	insects	44
**	Rhinolophoidea			
*	Megadermatidae (False vampire bats)	Old World Tropics	animals from insects to vertebrates	5
*	Nycteridae (Slit-faced bats)	Africa to Java and Sumatra	from insects to vertebrates	13
*	Rhinolophidae	Old World	insects	69
*	Hipposideridae (Old world leaf-nosed bats)	Old World Tropics	insects	56

Echolocation of bats is a mode of detecting obstacles, by emitting high frequency ultrasounds and analysing the echoes that hit and come back from the obstacles.

megachiroptera represents large bats, in which 150 species are included. They are characterized by large eyes and small and simple ears with no echolocation ability. *Microchiroptera* consists of about 700 species. They are smaller in size, have small eyes and the ability

*** Suborder ** Super family * Family	Distribution	Diet	No. of species
** Phyllostomoidea			
* Noctilionidae (Bulldog bats)	New World Tropics	insects and fish	2
* Mormoopidae (Mustached bats)	New World	insects	8
* Mystacinidae (Short-tailed bats)	New Zealand	insects, fruit, nectar, carrion	1
* Phyllostomidae (New world leaf- nosed bats)	New World Tropics	insects, fruit, pollen, verte- brate's blood	123
** Vespertilionoidea			
* Natalidae (Funnel-eared bats)	New World Tropics	insects	4
* Furipteridae (Thumbless bats)	New World Tropics	insects	2
* Thyropteridae (New world disk- winged bats)	New World Tropics	insects	2
* Vespertilionidae (Plain-nosed bats)	Worldwide	insects, fish and other vertebrates	283
* Myzopodidae (Old world disk- winged bats)	Madagascar	insects	1
* Molossidae (Free-tailed bats)	Tropics	insects	82

to echolocate. Their ears are larger in size with a variety of elaborate flaps and folds surrounding the ear canal in addition to the pinna. One common structure seen here is the tragus. It is a spear shaped part projecting upwards from the base of the ear up till the middle
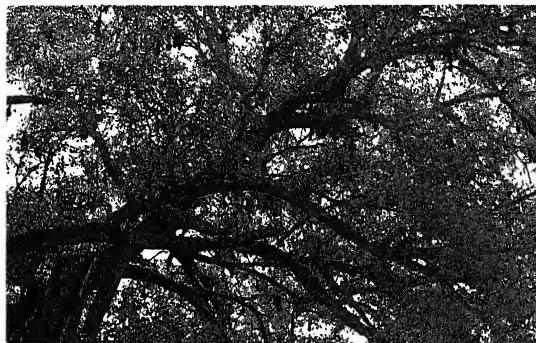
Figure 1(a) (top left) A colony of the Indian flying fox Pteropus giganteus *roosting during daytime hanging from a banyan tree.*
Figure 1(b) (top right) The 'flying fox', so called since the facial features appear like those of a fox.

Figure 2 (bottom left) A portion of a large colony of the fruit bat Rousettus leschenaulti, *occupying an unused temple.*

Figure 3 (bottom right) The short-nosed fruit bat Cynopterus sphinx.

portion of the ear canal. The brown long-eared bat of Europe has large pinnae that are almost as long as its body. A few microchiropterans have a fleshy appendage surrounding the nostrils (*Figures 4,5, and 11*) known as 'nose-leaf'. It aids in echolocation.

The size of the bats shows an amazing variety. The megachiropterans are commonly known as flying foxes. The Indian flying fox *Pteropus giganteus* is one of the largest bats in the world. It weighs about 1.5 kg. and has a wingspan of more than 1 metre. On the other hand the hog-nosed bat (microchiropteran) weighs only 2 g. and has a wingspan of about 13 cm. This is the smallest known mammal. It was discovered in 1973 in Thailand by a zoologist Kitti Thonglongyai and is hence named *Craseonycteris thonglongyai*.

Habitats

Bats live in different types of habitats. The Indian flying fox *Pteropus giganteus* lives in colonies and hangs from the branches of
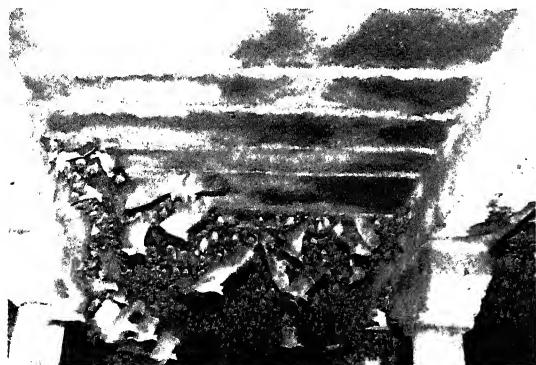
trees such as banyan, mango and tamarind. The majority of the microchiropterans live in dark caves and rocky crevices. Other places like temples, hollow trees, culverts, the underside of bridges and unused buildings are also occupied by bats. A few species modify large palm leaves by biting a series of holes across the centre of the leaf which makes the leaf hang like a tent. The bats hang from the apex underneath. The short-nosed fruit bat of India, *Cynopterus sphinx* modifies the twigs of the creeper plant *Vernonia*, shapes it into a dome-like tent and lives inside.

Figure 4 The leaf-nosed bat Hipposideros speoris.

Food

Regarding the type of food, the megachiropterans feed mainly on fruits like grapes, guavas, custard apples, bananas, papayas and chickoos. In addition they feed on leaves, petals and nectar. The microchiropterans generally feed on insects like moths, beetles, crickets and mosquitoes. In addition, a few species of bats feed on small animals like fish, frogs, mice and geckos. The vampire bats, which live only in Central and South America, feed upon the blood of large animals like horses, cows and pigs. Each night a vampire consumes 10-15 ml. of blood.

In India, of the 12 species of fruit bats found, three are very common: the Indian flying fox *Pteropus giganteus*, the fulvous fruit bat *Rousettus leschenaulti* and the short-nosed fruit bat *Cynopterus sphinx*.

Bats in India

In India, of the 12 species of fruit bats found, three are very common: the Indian flying fox *Pteropus giganteus*, the fulvous fruit bat *Rousettus leschenaulti* and the short-nosed fruit bat *Cynopterus sphinx*. About 100 species of microchiropteran bats live in different parts of India including the Eastern Ghats, Western Ghats, Himalayas and Andaman and Nicobar islands. A few species which live in South India are illustrated in the following section.

Figure 5 The leaf-nosed bat H. fulvus.

• The Indian flying fox (*Pteropus giganteus*) as mentioned earlier lives in a colony consisting of hundreds of individuals (*Figure 1a, b*). They roost only in the branches of trees. Compared to other bats, *Pteropus* has fewer predators. Humans are potential predators and may hunt this species as a source of protein. However, we once
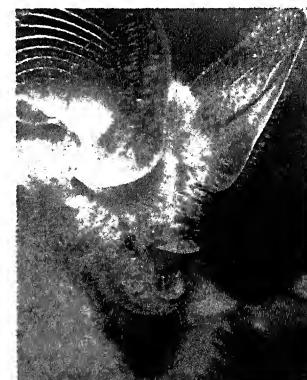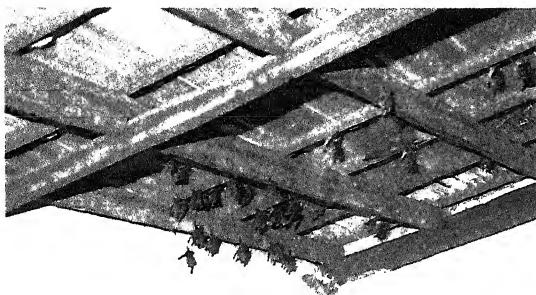
Figure 6 (top left) A portion of the colony of the leaf-nosed bat H. ater, occupying an unused building.

Figure 7 (top right) The sac-winged bat Taphozous nudiventris kachhensis.

Figure 8 (bottom left) An adult male tomb bat T. melanopogon has black fur, called a 'beard' at the ventral neck area.

Figure 9 (bottom right) The free-tailed bat Tadarida aegyptiaca.

observed a python capturing and feeding on this large-sized bat.

● The fulvous fruit bat (*Rousettus leschenaulti*) roosts in clusters in unused buildings and temples. It weighs about 110 g. *Figure 2* shows a section of a large colony of about 15,000 individuals occupying an unused temple.

● The short-nosed fruit bat *Cynopterus sphinx* (*Figure 3*) prefers foliage roosting. It roosts in trees under the dried leaves of Kitul palm, creeper plant *Vernonia*, etc. It weighs about 45-50 g. It lives in clusters of small colonies of about 10-30 individuals. The adult males are yellowish in colour.

The following are the echolocating and insect-eating bats:

● The leaf-nosed bat *Hipposideros speoris* (*Figure 4*) lives in colonies in caves, temples, unused buildings, etc. It weighs about 11 g. The fur of the adults is yellow. It breeds all through the year but more young are born during September-October. The individuals do not cluster together.
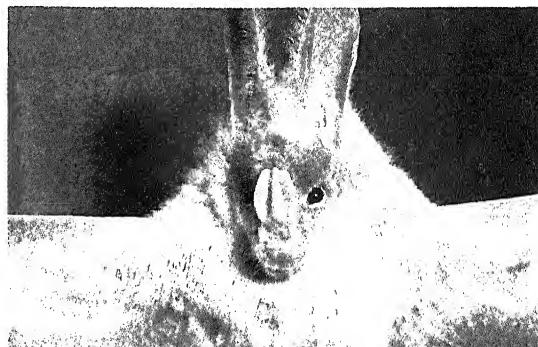
The leaf-nosed bat *H. fulvus* (*Figure 5*) lives in colonies mainly in caves. It weighs about 9 g. The fur of the adults is white on the ventral and yellow on the dorsal sides. They breed from April to July.

The leaf-nosed bat *H. ater* prefers unused buildings (*Figure 6*) for roosting during day time. It weighs about 7g. It breeds from September to December. A unique feature of the hipposiderid bats is that they frequently move their head and ears while roosting.

The sac-winged bat *Taphozous nudiventris kachhensis* (*Figure 7*) lives in vertical rock crevices and weighs about 40 g. Upon close human approach the individuals crawl on all fours and move into deeper parts of the crevice. These bats breed during October-November.

The tomb bat *Taphozhous melanopogon* (*Figure 8*) lives in caves and temples and weighs about 25 g. The adult males have well developed black beards. They breed during April-May.

The free-tailed bat *Tadarida aegyptiaca* (*Figure 9*) also lives in narrow rock crevices. It has wrinkled lips especially on the upper jaw. It weighs about 20 g. It has a short tail extending from the interfemoral membrane. The individuals in a colony are noisy even during the day. These bats breed during September.

The mouse-tailed bat *Rhinopoma hardwickei* (*Figure 10*) lives in caves and wide rock-crevices. It weighs about 17 g. It has a long and slender tail. It breeds twice a year in April and November.

Figure 10 (top left) A colony of mouse-tailed bats **Rhinopoma hardwickei,** *occupying an unused tunnel in a building.*

Figure 11 (top right) The Indian false vampire bat **Megaderma lyra.**

● The Indian pygmy bat *Pipistrellus mimus* is the smallest bat in India, weighing about 3 g. It lives in all sorts of places with narrow cracks and crevices. It slips into these spots on its back and the underparts of its body which come in contact with stones or wood. Twin babies is a characteristic of this bat. An individual female gives birth to twins three times a year thus producing six infants in a year. This is a fast rate of reproduction among bats.

● The Indian false vampire bat *Megaderma lyra* (*Figure 11*) lives in caves and unused buildings. It weighs about 40 g. It is carnivorous — feeds upon frogs, mice, geckos, larger insects, etc. It has large and medially fused pinnae which receive the weak sound created by the movement of its prey on the ground. It breeds from March to June.

Some of the species of bats mentioned above live in different parts of the same caves or buildings.

Address for correspondence
G Marimuthu
Department of Animal
Behaviour & Physiology,
School of Biological Sciences,
Madurai Kamaraj University,
Madurai 625 021. India.

Suggested Reading

A Brosset. Bats of Central and Western India. *J. Bombay Nat. His. Soc.* 59 1-57, 583-624, 707-746. 1962.
M B Fenton. Bats. Facts On File. New York, Oxford. 1992.
H Khajuria. Records of the Zoological Survey of India, Studies of Bats. Part I. Occasional Paper 13, 1-59. 1979.
H Khajuria. Taxonomical and Ecological Studies on the Bats of Jabalpur District, MP, India. Part II. Occasional Paper 19, 1-69. 1980.
W Schober. The Lives of Bats. Croom Helm, London and Canberra. 1984.

And the latest from the world of honey-bees and their dance language.

Breakdance ? That's bad language !

MOHAN DEVADAS
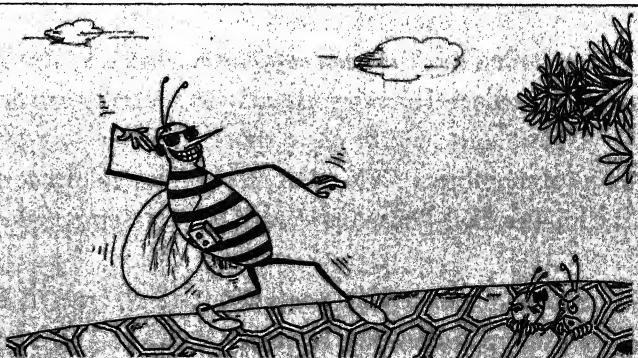
Molecule of the Month

A 'Hexacationic' Benzene Derivative!

Uday Maitra

A stable hexacationic benzene derivative has been prepared by the displacement of all the fluorine atoms of hexafluorobenzene by six molecules of 4-dimethylamino-pyridine.

Uday Maitra is a member of the Organic Chemistry Faculty at Indian Institute of Science.

The nucleophilic displacement of a halogen atom attached to an aromatic ring is not a very favourable process.[1] Polyfluorinated aromatics are interesting in this regard, since the fluorine atom can act as a leaving group, as well as an activating group. Hexafluorobenzene, therefore, has been a popular molecule to examine the displacement of all the fluorine atoms by nucleophiles.

Imagine the (hypothetical) displacement of the fluorine atom of fluorobenzene by 4-dimethylaminopyridine (DMAP). The product of this reaction is a cation (an *onium salt*), which of course is resonance stabilized as shown in *Figure 1*. If we try to extend this idea with hexafluorobenzene, we would expect a hexa(onio)substituted benzene! At first sight this would appear to be impossible because of the accumulation of like charges. However, German chemists Robert Weiss and coworkers have

[1] Usually, aromatic compounds undergo *electrophilic* substitution reactions. Nucleophilic substitution on an aromatic ring requires electron withdrawing groups on the ring (or, special reaction conditions, as in a benzyne mechanism). Such 'activated' aromatic halides undergo nucleophilic substitution by an addition-elimination sequence. The negatively charged addition product is called a *Meisenheimer complex*.
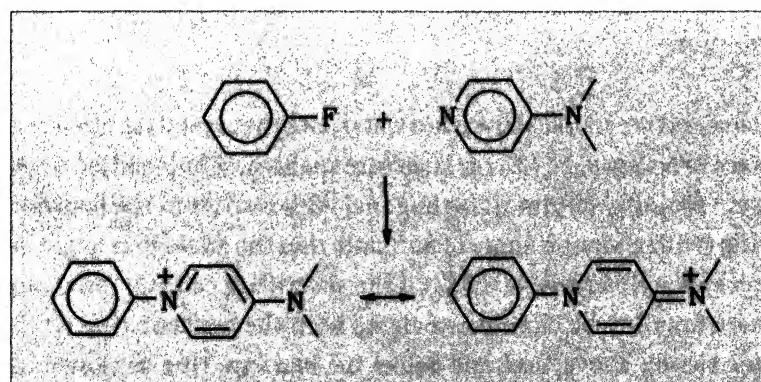

Figure 1 Resonance stabilization in N-phenyl-4-dimethylaminopyridinium salt.
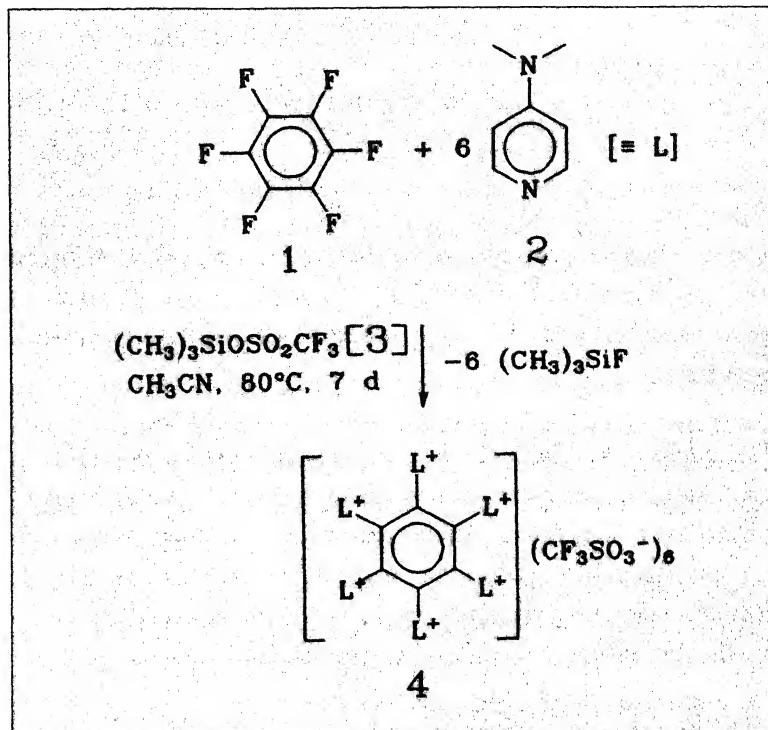
recently shown that it is indeed possible (R. Weiss *et al.*, *Angew. Chem. Int. Ed. Engl.*, 1995, *34*, 1319). They simply reacted hexafluorobenzene (1) and DMAP (2) with trimethylsilyl trifluoromethanesulfonate (TMS-triflate, 3) in refluxing acetonitrile for seven days. What was formed in almost quantitative yield is the per(onio)² product 4, with six triflate counter ions (*Figure 2*). Clearly, the formation of this product is aided by the formation of strong Si-F bonds in the other product of the reaction.

Salt 4 was crystallized from hot water, and the crystals so obtained were investigated by X-ray structure analysis. The pyridine rings were found to be almost orthogonal with respect to the benzene ring (with a torsion angle of 80°; note that the analogous angle in hexaphenylbenzene is 65°). Of the six triflate counter ions, two were found to be closely associated with the benzene ring (with one sitting above, and one below the benzene ring as shown in
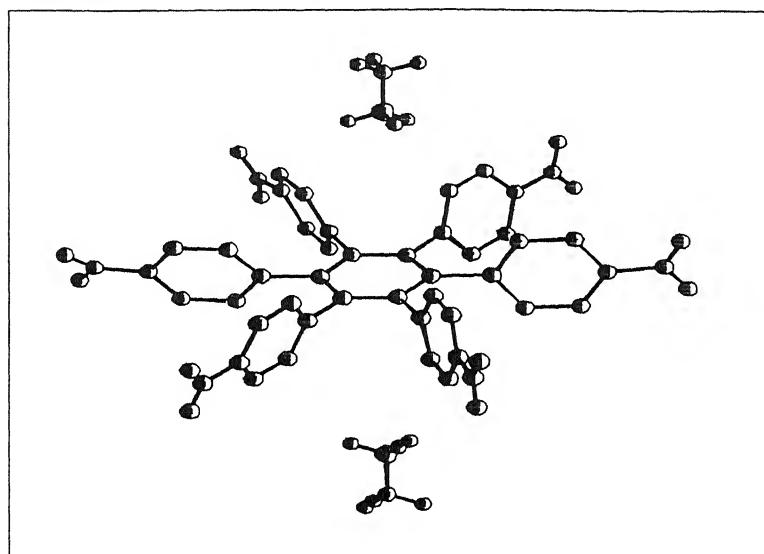
Figure 3).

How would you expect the reactivity of **4** to be? Since it is substituted by six cationic units, one would certainly expect greatly facilitated nucleophilic substitutions. In fact, if **4** is simply boiled with dilute $NaHCO_3$ solution for 30 minutes, compound **5**, resulting from the displacement of one of the DMAP units, is formed (*Figure 4*). It is interesting to know that hexacyanobenzene also undergoes an analogous reaction under comparable conditions. The difference between **4** and hexacyanobenzene is that in **4** the activation is exclusively via electrostatic effects, whereas in hexacyanobenzene -*M* effect[3] plays an important role.
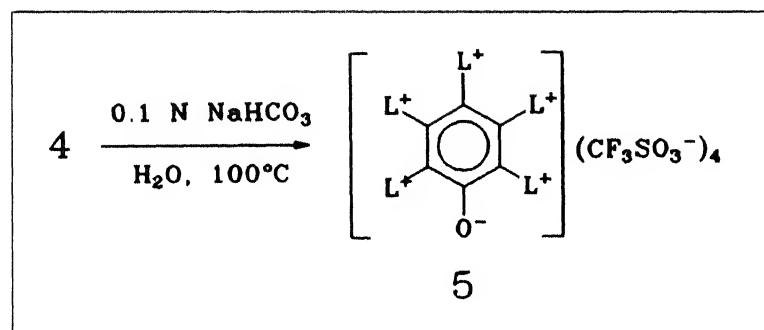
[3] The -*M* effect refers to electron withdrawing ability through resonance (mesomeric effect). Electron donation through resonance is represented by +*M*.

Address for correspondence
Uday Maitra
Department of
Organic Chemistry,
Indian Institute of Science,
Bangalore 560 012, India.

Figure 4 Facile nucleophilic displacement of one ligand from 4.

Classroom

In this section of Resonance, *we invite readers to pose questions likely to be raised in a classroom situation. We may suggest strategies for dealing with them, or invite responses, or both. "Classroom" is equally a forum for raising broader issues and sharing personal experiences and viewpoints on matters related to teaching and learning science.*

? An observer points a torch at a mirror which is moving away at a velocity v. The frequency of the light is ν_0. The light reflected at normal incidence has a lower frequency ν. A calculation including special relativity, gives the result

$$\nu = \nu_0 \left(\frac{c - v}{c + v} \right)$$

where c is the speed of light. Why is this different from the standard Doppler shift formula for a source moving at a velocity of v? This reads

$$\nu = \nu_0 \left(\frac{c - v}{c + v} \right)^{1/2}$$

Can one speak of a velocity of the image and if so what is it?

An unusual case of suicide by drowning, boys ...

MOHAN DEVADAS
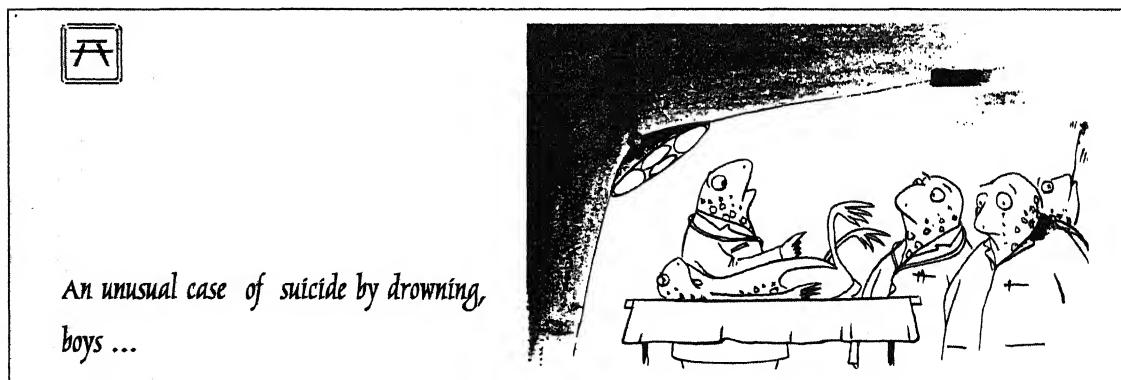
Think It Over

This section of Resonance *is meant to raise thought-provoking, interesting, or just plain brain-teasing questions every month, and discuss answers a few months later. Readers are welcome to send in suggestions for such questions, solutions to questions already posed, comments on the solutions discussed in the journal, etc. to* Resonance, *Indian Academy of Sciences, Bangalore 560 080, with "Think It Over" written on the cover or card to help us sort the correspondence. Due to limitations of space, it may not be possible to use all the material received. However, the coordinators of this section (currently A Sitaram and R Nityananda) will try and select items which best illustrate various ideas and concepts, for inclusion in this section.*

Hydrocarbon Acts Funny!

Photon Rao, Department of Organic Chemistry, Indian Institute of Science.

Given a hydrocarbon of formula $C_{115}H_{90}$ the last thing one expects is that it is ionic. But, yes, that is what it was reported by the Japanese chemists K Okamoto, T Kitagawa, K Takeuchi, K Komatsu and K Takahashi (*J. Chem. Soc., Chem. Commun.*, **1985**, 173). When tris (5-isopropyl-3,8-dimethylazulenyl) cyclopropenylium perchlorate, $C_{48}H_{51}{}^{+} ClO_4{}^{-}$ and potassium tris (7-H-dibenzo [c,g] flurenylidenemethyl), $K^{+}C_{67}H_{39}{}^{-}$ were mixed in tetrahydrofuran the organic solid **A** was obtained as greenish black crystals — stable upto six months under argon. Even though purely organic examples of ionic compounds were known in solution, this was the *first hydrocarbon* to earn such a distinction. Usually, carbocations and carbanions, if mixed together give rise to a **C-C** single bond. Can you figure out the tricks employed by the researchers to realise the hydrocarbon-only ionic compound?

A

The 1995 Nobel Prize in Physiology or Medicine

Flies Take Off at Last!

Vidyanand Nanjundiah

The Nobel Prize in Biology (correctly stated, in Physiology or Medicine) for 1995 was awarded jointly to Edward B Lewis (77 years old) of the California Institute of Technology, Christiane Nüsslein-Volhard (52) of the Max Planck Institute for Developmental Biology in Tübingen and Eric Wieschaus (48) of Princeton University. Among other reasons given in the citation, the award was ".. for discovering how genes control the early structural development of the body ".

These three workers studied the genetic basis of development — meaning the set of processes that convert an egg to an adult — in the fruit fly *Drosophila melanogaster*. They asked, in what manner do genes contribute to the changes that take place as the fly's egg first becomes a larva and eventually an adult?

Specifically, are there genes that guide the anterior-to-posterior patterning of the fly's body into head, three distinct thoracic segments and eight abdominal segments? The answer is: yes, there are such genes, and an unexpectedly small number at that. Their existence, as well as the roles they play, can be inferred from what happens in mutant flies lacking one or the other of these genes.

To start with, let us recall that the fly is an insect. Also, it is a member of a larger group, the arthropods, that are characterised by a basic body plan that consists of paired repeated segments called metameres. Metameric design is obvious when we look at the larval worm-like stage of the fly, but close observation shows that it also exists in the adult. In other words larval segments resemble each other markedly, but the segments in the adult are very different from one another. In effect, the question asked by all three prize winners was, why are all segments not the same?

E B Lewis started his work in the 1940's with a mutant, known as *bithorax*, that had been
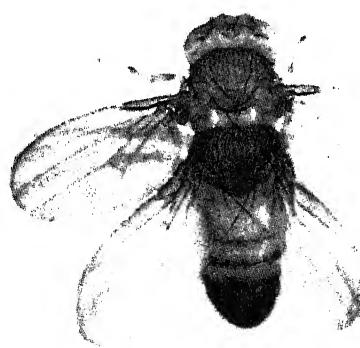
Figure 1 A typical four-winged fly generated when certain **bithorax** *mutant alleles are brought together in heterozygous combinations (Reprinted from* **Current Science.** *69:799. 1995)*

isolated earlier by the distinguished geneticist, C B Bridges. Bridges had found that flies with two copies of the mutant bithorax gene (*bx* / *bx*) tended to develop portions of an extra pair of wings. In normal flies the second segment of the thorax carries a pair of wings while the third thoracic segment has a pair of balancers, or halteres, attached to it. *Bithorax* flies have the normal pair of wings all right; but in addition, the front half of each haltere is replaced by a half-wing. The transformation stands out because halteres are tiny compared to wings and the juxtaposition of the mismatched halves presents a strange sight (*Fig-*

ure 1). Such a phenomenon, where a segment, or a portion of a segment, gets replaced by another segment, is known as homeosis.

Lewis identified a whole series of homeotic genes in what came to be known as the *Bithorax complex*. A number of important conclusions emerged from Lewis's study of homeotic mutants. Let us denote the segments of the body by the symbols H (for head), T_1, T_2, T_3 (for the three segments of the thorax) and A_1, A_2, ... A_8 (for the eight segments of the abdomen; see *Figure 2*). First of all, mutations in two genes, *bx* and *pbx*, cause a $T_3 \longrightarrow T_2$ switch. Therefore, in the normal fly, the genes in question must be needed to make T_3 develop differently from T_2. One says that the action of *bx* and *pbx* confers segmental identity to T_3. The same thing can be said in different words: the wild-type gene breaks an underlying sym-
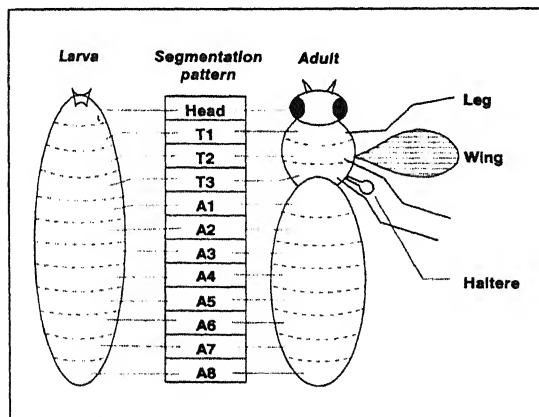

Figure 2 A diagrammatic representation of segmentation in Drosophila. The worm-like pattern seen in the larva is retained in the adult fly. T1-T3 are thoracic segments, A1 to A8 are abdominal segments. Drawn using a computer (based on Gilbert 1988).

metry between T_2 and T_3. On the other hand, the double $bx\,pbx$ mutant reveals the existence of the symmetry. Similarly, other genes of the *Bithorax complex* are responsible for breaking the symmetry between A_1 and T_3, between A_2 and A_1, and so on. The deepest symmetry of all is unveiled when the entire complex is deleted, because in that condition all segments posterior to T_2 look like T_2. As Lewis pointed out, both single gene mutations and the deletion of the complex appeared to evoke traces of the fly's evolutionary ancestors, four-winged insects like dragonflies and worms. However, the mutants are not *true* throwbacks to an ancestral form; the four-winged fly cannot fly and the $H\text{-}T_1\text{-}T_2\text{-}T_2\text{-}\ldots\text{-}T_2$ larva dies very early. This means that other genes must also have evolved in the course of the evolution from worms to flies. As you must have noticed, the *Bithorax complex* does not seem to be required in H, T_1 and T_2. It turns out that these segments depend on the activity of another set of homeotic genes, also constituting a complex, known as *Antennapedia*. Genes such as those of the *Bithorax complex* have been called master genes or control genes for the specification of body pattern.

Nüsslein-Volhard and Wieschaus are responsible for identifying three families of other master genes which regulate body pattern well before the *Bithorax complex* becomes active.

(They do so soon after fertilization, and in some cases even before the embryo has become subdivided into cells.) The two of them treated fly embryos with a chemical mutagen and with the aid of standard but extremely tedious procedures, set about looking for as many genes as possible that could influence body patterning along the anterior-posterior axis. The results were astonishing in two respects. Firstly, the total number of candidate genes turned out to be unexpectedly small, just 15 in all. Today, 16 years later, the number still remains small. (It would not have been considered strange if hundreds of genes had participated in the major decision-making steps necessary for the specification of the body plan.) Secondly, the genes fell naturally into three families. Within each family, mutations had striking but distinct effects.

● The first class of genes were named *gap* genes. When mutated so that their function was lost, they gave rise to larvae with gaps of varying extent in the segmental pattern.

● Next were the *pair-rule* genes, and they caused the most surprise when people first heard about them. Mutations in these genes cause the elimination of portions of the body pattern in a periodic fashion. The strange thing was that the period did not correspond

to the length of one segment, as might have been expected, but to that of *two* segments. By skipping alternate segments, the pair-rule genes point to an underlying periodicity in the body plan with two segments as the unit.

● Finally, there were the *segment polarity* genes. When mutated they led to the disappearance of a portion of each segment and its replacement by the remaining portion. However, the replacement has its polarity inverted, meaning that the duplicated portion is a mirror image replica of the undisturbed part.

This report of Nüsslein-Volhard and Wieschaus was so illuminating that it led to an explosive burst of activity on the part of researchers all over the world. Thanks to it we can today begin to build a model of how genes specify the body plan of *Drosophila*. The basic idea seems to be that there is a hierarchical order to genetic activity. Genes that are higher in the hierarchy specify gross features of the body plan and genes that are lower down in the hierarchy sharpen the specification further.

What I have described is the barest outline of what we know about how genes regulate the fly's body pattern. Even so, it is apparent that the picture we are beginning to glimpse is one of both order and complexity. Further research is going on in an attempt to decipher

> The achievements of Lewis, Nüsslein-Volhard and Wieschaus constitute a striking vindication of the power of formal genetic analysis in the study of development. "Look for the genes behind the phenomenon" has turned out be a successful philosophy.

the details, especially the molecular details, of the working of these 'master' genes. The achievements of Lewis, Nüsslein-Volhard and Wieschaus constitute a striking vindication of the power of formal genetic analysis in the study of development. "Look for the genes behind the phenomenon" has turned out be a successful philosophy. Their work needed very little by way of sophisticated equipment; perhaps a good dissecting microscope, but that is all. Why then did no one attempt it earlier?

Suggested Reading

S F Gilbert. Developmental Biology. (Second Edition) Sinauer Associates Inc., Sunderland, Massachussetts. 1988.

E B Lewis. A Gene Complex Controlling Segmentation in *Drosophila*. *Nature*. 276:565-570. 1978.

C Nusslein-Volhard, E Wieschaus. Mutations Affecting Segment Number and Polarity in *Drosophila*. *Nature*. 287:795-801. 1980.

P A Lawrence. The Making of a Fly - The Genetics of Animal Design. Blackwell Scientific Publications, Oxford. 1992.

Vidyanand Nanjundiah is at the Developmental Biology and Genetics Laboratory, Indian Institute of Science, Bangalore 560 012. India.

> Genes such as those of the *Bithorax complex* have been called master genes or control genes for the specification of body pattern.

An Important Mathematical Theory Explained

A Well Written Book for the B.Sc. and M.Sc. Student

S Thangavelu
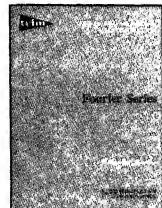
Fourier Series.
Rajendra Bhatia
Hindustan Book Agency, Delhi. 1993
pp.106. Rs.95.

> The book abounds in exercises and has a nice appendix on the historical development of the subject.

The theory of Fourier series is one of the most powerful theories that mathematicians have developed. It has a wide range of applications, not only in mathematics and natural sciences, but also in engineering fields. Any student of science and engineering must learn the rudiments of the theory. Although there are several excellent text books on the subject they are of no use to poor Indian students studying in remote colleges without access to good libraries. It is a pity that in a vast country like ours, good text books are not written and published at affordable prices. The book under review, a small book on Fourier series, is the outcome of an attempt to alleviate the scarcity of reasonable mathematical text books in the country.

A most books on Fourier series would do, this one introduces the subject through a partial differential equation, in the present case, the equation of Laplace. Once Fourier series is defined, one can take off and develop the subject in several directions. As this book is mainly aimed at B.Sc. final year students or M.Sc. first year students without any knowledge of functional analysis, only very basic material like pointwise convergence and convergence in L^2 and L^1 are treated. To exemplify the power of Fourier series the author has included some results from ergodic theory and number theory, a discussion on the isoperimetric problem and of course the solution to the wave equation. The book abounds in exercises and has a nice appendix on the historical development of the subject.

There are five chapters in this book. In the first chapter the author treats the Laplace equation $\partial_x^2 u + \partial_y^2 u = 0$ in the unit disc in the complex plane. He solves this equation in polar coordinates and in the process defines the Fourier series of a periodic function. Chapter 2 deals with convergence of Fourier series; the author treats Abel and Cesaro summability, and proves the celebrated theorem of Fejer regarding the pointwise convergence of the Cesaro means of Fourier series of continuous functions. The third chapter deals with sine and cosine series, functions with arbitrary periods and the Gibbs phenomenon. The L^2 convergence of Fourier series is considered in chapter 4 and, as already mentioned, the last chapter deals with some applications.

This is a very well written book and can be used effectively at the B.Sc. final year and

M.Sc. first year levels. The author has done justice to the good cause of writing text books for Indian students.

S Thangavelu is with Statistics and Mathematics Unit, Indian Statistical Institute, R V College Post, Bangalore 560 059.

A Haldane Primer: Getting Messages Across

"A Box of Reusable Chocolates that can be Savoured Afresh Each Time"

S Krishnaswamy

Haldane was a scientist known for his contri-
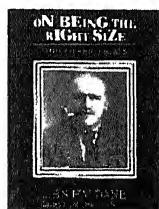
On Being the Right Size and Other Essays
J B S Haldane
Oxford University Press, Delhi. 1992.
pp. 187. Rs. 90.

bution to the theory of evolution. He and others developed the theory of population genetics. He was a liberal individualist, a leading communist who contributed a weekly article to the *Daily Worker*. He was supreme as a popularizer of science, because he saw connections that others missed. This book is a collection of his essays, compiled by John Maynard Smith who in his introduction mentions their impact on him when he came across them in school. Maynard Smith says " fifteen years later, when I decided to leave engineering and train as a biologist, I entered University College, London, where Haldane was at that time a professor, and became his student and later his colleague."

The essays covered in the book range from articles intended for a scientific readership to others written for a daily newspaper, and are comprehensible without the need for any special technical knowledge. The book is inexpensive enough to make it affordable and worthy enough to make efforts to locate and buy it. The introduction by Maynard Smith is evocative and gives an endearing perspective of Haldane and his essays. There are in all nineteen essays here, with three small bits in the appendix. The essays are arranged in chronological order and cover the period from the 1920s to the 1940s. The articles: 'Is History a Fraud?', 'God-Makers', 'The Origin of Life', ' What "Hot" Means' and 'Cats' appeared in the *Daily Worker* . Most of the other essays

This book is a collection of his essays, compiled by John Maynard Smith who in his introduction mentions their impact on him when he came across them in school.

J B S Haldane (1892-1964)

Haldane is today remembered mainly for his work on evolution, but that is only because his first rate work in several other areas pales in comparison with his work on evolution.

Haldane's contributions to developing a quantitative theory of evolution, or population genetics, were fundamental. The gigantic task of proving beyond all doubt that Mendelian principles could provide the link between genetics and evolution as demanded by natural selection (as Darwin's theory was called) was accomplished by J B S Haldane, R A Fisher, S Wright and S S Chetverikov.

Haldane's interest in genetic structure was lifelong. At the age of 16, he discovered the phenomenon of *linkage* (in mice), which means that genes are not free to 'move' independent of one another. He was also the first to discover linkage in humans; he pre-sented the first genetic map of a human chromosome and the first estimate of a human mutation rate. Realizing the importance of Garrod's studies on the inborn errors of metabolism, he considered the 'one gene one enzyme' hypothesis well before Beadle and Tatum.

A general sympathy for the underdog, the attraction of socialist ideals and disgust for the workings of capitalism made him a marxist. He spent the last seven years of his life in India working initially at Indian Statistical Institute, Calcutta, and later at the Genetics and Biometry Laboratory in Orissa. His behaviour sometimes seemed eccentric but behind his rough exterior was an extremely charming and kind man, full of humour, a truly humble person with the curiosity which is the hallmark of most great scientists.

(Courtesy V Nanjundiah, adapted from Current Science)

---

have appeared earlier in the collections *'Possible Worlds and Other Essays'* and *'A Banned Broadcast and Other Essays'*. The articles that featured in the *Daily Worker* differ stylistically from the other essays and are aimed at a readership with less formal education. However his perspectives on the science-society relation remain the same as in the other articles.

A few words from Maynard Smith will emphasize how Haldane's articles are different from regular popular science articles. "It was characteristic of him (Haldane) to use popular articles to propose original ideas. 'The Origin of Life' is perhaps the most important. The ideas presented here contributed significantly to the fact that the topic is now one for experimental study and not for philosophical speculation...I do not think that anyone today is writing scientific articles for the daily press that equal these in scientific content or entertainment value. To illustrate their scientific content, the article 'Beyond Darwin' contains a particularly clear account of Darwin's idea about the relationship between sexual dimorphism and polygyny, which has since become a popular topic of research. I was amused to

find that the same article contains a false argument about species which destroy their food supply and starve to death. I learnt the fallacy of such 'group-selectionist' arguments from Haldane himself when I was his student."

The article on 'How to write a Popular Scientific Article' that is part of this collection gives a glimpse of how Haldane wrote essays whose general sense of argument was so clear. In very conversational terms he builds upon the writing of popular science articles and leads you on to the different aspects such as the audience, the subject matter, the place where you want to publish it, the depth to which it must be written, the language to be used, etc. The essay is actually aimed at science practitioners and aims to help them convey their work to the larger public. Some excerpts will help you get the general trend. "So far you have probably written two main types of articles. Firstly, answers to examination questions in which you tried to show how much you knew about some topic. And secondly, scientific papers or technical reports which dealt very exhaustively with a small point. Now you have to do something different. You are not trying to show off; nor are you aiming at such accuracy that your readers will be able to carry out some operation. You want to interest or even excite them, but not to give them complete information. You must therefore know a very great deal more about your subject than you put on paper. Out of this you must choose the items which will make a coherent story." He then goes on to give an

example of a skeleton of an article on cheese to illustrate his points.

Another article I found very entertaining was the one on 'Cats'. He starts off with a simple observation that the number of cats in Britain is not known unlike the number of adult dogs which are kept track of since they are taxed. He builds up to talk about how there are no peculiar shapes or sizes in cats. He goes on to compare the social and breeding behaviours of cats and humans. In this manner we are introduced to how different animals have their sensory areas on the body mapped on to the brain. It concludes with "A cat or dog can be gentle with its whole body. So dogs and cats can play with us, and we with them. In fact they play with children very much as equals, and quite understand that they must not use their full strength."

The title essay "On Being the Right Size" is a gem and is reminiscent of d'Arcy Thomson's *On Growth and Form*. He shows elegantly the connection between simple geometry, physics and biological forms. He throws light on why it is that if one were to drop a mouse down a thousand-yard mine shaft, on reaching the bottom it gets a slight shock and walks away. But if you did the same with a rat, a man and a horse in that order, they are respectively killed, broken or splashed. And why the eyes

> The title essay "On Being the Right Size" is a gem and is reminiscent of d'Arcy Thomson's *On Growth and Form*.

of small animals have to be in much larger proportion to their bodies than our own, while large animals only require relatively small eyes — those of the whale and elephant being little larger than our own. The message is that, for every type of animal there is an optimum size. In his characteristic manner he makes the connections with human societies: " And just as there is a best size for every animal, the same is true for every human institution... To the biologist the problem of socialism appears largely as a problem of size... But while nationalisation of certain industries is an obvious possibility in the largest of states, I find it no easier to picture a completely socialized British Empire or United States than an elephant turning somersaults or a hippopotamus jumping a hedge."

The final bits in the appendix again prove instructive with the explanation by Maynard Smith and his account of what has happened since then to Haldane's ideas. Maynard Smith says: "Scattered through Haldane's writings — both the popular articles and scientific papers — there are hints of things to come. He sketches some experiment, or theory, or field of investigation which later, in other hands, has become important. There are many possible reasons why he failed to follow up these hints himself: he was too impatient to be good at raising grants to support experimental work, the techniques needed to test an idea were not yet developed, or he simply had too many other things to think about."

In all, a box of reusable chocolates that can be savoured afresh every time with relish and pleasure. If you are not one already you will turn a Haldane fan.

S Krishnaswamy is with the Bioinformatics Centre, School of Biotechnology, Madurai Kamaraj University, Madurai 625 021.
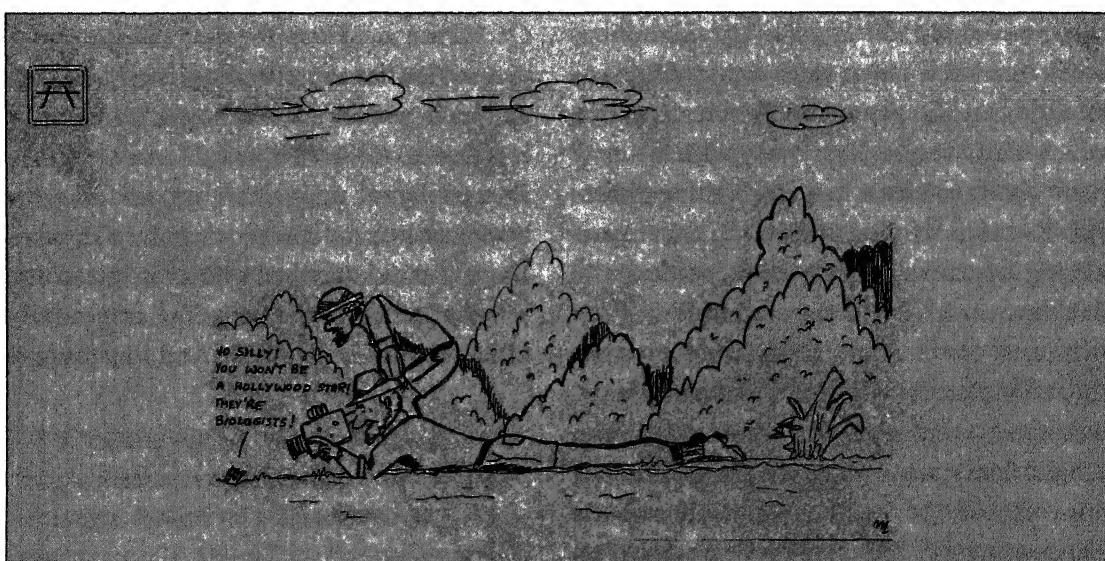
Books Received

Functional Analysis P K Jain, O P Ahuja, Khalil Ahmad New Age International Pvt. Ltd. Rs.130.	*Srinivas Ramanujam (in Kannada)* S Balachandra Rao Kannada Pustaka Pradikara Rs.15.	*Indian Mathematics and Astronomy.* *Some Landmarks* S Balachandra Rao Janana Deep Publications. Bangalore Rs.75.

Information and Announcements

Academy Initiative in University Education

The Indian Academy of Sciences, as a part of its initiative in university education, has limited funds available to support:

- Short visits (4-8 weeks) by college/university students and teachers to Academy Fellows for joint work on specific projects.
- Short intensive lecture series by Fellows in colleges and universities.

Proposals are invited by the Education Panel of the Academy (R Narasimha, Chairman; V S Borkar, Secretary) for the above from college and university teachers and students. They should include a brief resume, a one page description of the proposed activity including its duration and tentative dates, and a letter of consent from the concerned Fellow, and should be mailed to: *Executive Secretary, Indian Academy of Sciences, PO Box 8005, C V Raman Avenue, Bangalore 560080.*

Guidelines for Authors

Resonance - journal of science education is primarily targeted to undergraduate students and teachers. The journal invites contributions in various branches of science and emphasizes a lucid style that will attract readers from diverse backgrounds. A helpful general rule is that at least the first one third of the article should be readily understood by a general audience.

Articles on topics in the undergraduate curriculum, especially those which students often consider difficult to understand, new classroom experiments, emerging techniques and ideas and innovative procedures for teaching specific concepts are particularly welcome. The submitted contributions should not have appeared elsewhere.

Manuscripts should be submitted in *duplicate* to any of the editors. Authors having access to a PC are encouraged to submit an ASCII version on a floppy diskette. If necessary the editors may edit the manuscript substantially in order to maintain uniformity of presentation and to enhance readability. Illustrations and other material if reproduced, must be properly credited; it is the author's responsibility to obtain permission of reproduction (copies of letters of permission should be sent). In case of difficulty, please contact the editors.

Title Authors are encouraged to provide a 4-7 word title and a 4-10 word sub-title. One of these should be a precise technical description of the contents of the article, while the other must attract the general readers' attention.

Author(s) The author's name and mailing address should be provided. A photograph and a brief (in less than 100 words) biographical sketch may be added. Inclusion of phone and fax numbers and e-mail address would help in expediting the processing of manuscripts.

Summary and Brief Provide a 2 to 4 sentence summary, and preferably a one sentence brief for the contents page.

Style and Contents Use simple English. Keep the sentences short. Break up the text into logical units, with readily understandable headings for each. Do not use multiple sub sections. Articles should generally be 1000-2000 words long.

ations Use figures, charts and schemes liberally. A few colour illustrations may be
. Try to use good quality computer generated images, with neatly labelled axes, clear
. fonts and shades. Figure captions must be written with care and in some detail. Key
≥s of the illustration may be pointed out in the caption.

r Highlights, summaries, biographical and historical notes and margin notes presented
⩘el different from the main body of the text and which nevertheless enhance the interest
⋂ain theme can be placed as boxed items. These would be printed in a different typeface.
⌐ boxed item should fit in a printed page and not exceed 250 words.

⩘sted Reading Avoid technical references. If some citations are necessary, mention
◻s part of the text. A list of suggested readings may be included at the end.

⊓ It is preferable to place all the boxes, illustrations and their captions after the main text
◻rticle. The suggested location of the boxes and figures in the printed version may be
◻ in the text. In the printed version, the main text will occupy two-thirds of each page. The
�⊓ing large margin space will be used to highlight the contents of key paragraphs, for
◻aptions, or perhaps even for small figures. The space is to be used imaginatively to draw
◻⋂ to the article. Although the editors will attempt to prepare these entries, authors are
r◻ged to make suitable suggestions and provide them as an annexure.

Reviews

⋉owing types of books will be reviewed : (1) text books in subjects of interest to the journal;
⋂eral books in science brought to the attention of students/teachers; (3) well-known
:s; (4) books on educational methods. Books reviewed should generally be affordable to
⋈ts/teachers (price range Rs.50 to 300).

·◻oks will get preference in review. A list of books received by the academy office will be
⋔ed among the editors who will then decide which ones are to be listed and which to be
'ed.

Acknowledgements

Resonance gratefully acknowledges the help received from the following individuals:

B Bagchi, L Geetha, Milind Kolatkar, Gurumurthi V Ramanan, Jayant Rao, A Sumana Rao, V Ravi, Joseph Samuel.

R B Woodward (1917-1979) was one of the giants of modern organic chemistry. Born in Boston on April 10, 1917, Woodward (RBW) obtained his B.S. (1936) and Ph.D. (1937) from MIT. It is interesting to read what J F Norris, MIT organic chemistry professor at that time, said about RBW: "We saw we had a person who possessed a very unusual mind in our midst. We wanted to let it function at its best. If red tape which was necessary for other less brilliant students had to go, we cut it. We did for Woodward what we have done for no other student in our department, for we had no student like him in the department."

During his almost four decade long career at Harvard University, he synthesized a large number of complex natural products: Cholesterol (1951), Strychnine (1954), Lysergic Acid (1954), Reserpine (1956), Chlorophyll-a (1960), Cephalosporin C (1965), Vitamin B_{12} (1976), Erythromycin (1980), etc. He was awarded the Chemistry Nobel Prize in 1965 "for his outstanding achievements in the art of organic synthesis".

In addition to his synthetic achievements, he is also widely known for a set of orbital symmetry rules he developed with Roald Hoffmann for pericyclic reactions ("Woodward-Hoffmann rules"). Woodward was a pioneer in using modern instrumentation in chemical research. The empirical rules he developed for determining structural details from UV-Visible spectroscopy are still in use ("Woodward-Fieser rules").

Woodward trained over 400 students and postdoctoral fellows during his career. J.-M. Lehn, whose article appears in this issue, is one of his illustrious coworkers.

Robert Burns Woodward

1917-1979